

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

**IMAGES ARE BEST AVAILABLE COPY.**

**As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problem Mailbox.**

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
10 January 2002 (10.01.2002)

PCT

(10) International Publication Number  
WO 02/02606 A2

(51) International Patent Classification<sup>7</sup>: C07K 14/295,  
C12N 15/31, A61K 39/118

(21) International Application Number: PCT/IB01/01445

(22) International Filing Date: 3 July 2001 (03.07.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Date:

0016363.4	3 July 2000 (03.07.2000)	GB
0017047.2	11 July 2000 (11.07.2000)	GB
0017983.8	21 July 2000 (21.07.2000)	GB
0019368.0	7 August 2000 (07.08.2000)	GB
0020440.4	18 August 2000 (18.08.2000)	GB
0022583.9	14 September 2000 (14.09.2000)	GB
0027549.5	10 November 2000 (10.11.2000)	GB
0031706.5	22 December 2000 (22.12.2000)	GB

(72) Inventors; and

(75) Inventors/Applicants (for US only): RATTI, Giulio [IT/IT]; Chiron S.p.A., Via Fiorentina, 1, I-53100 Siena (IT). GRANDI, Guido [IT/IT]; Chiron S.p.A., Via Fiorentina, 1, I-53100 Siena (IT).

(74) Agents: HALLYBONE, Huw, George et al.; Carpmaels & Ransford, 43 Bloomsbury Square, London WC1A 2RA (GB).

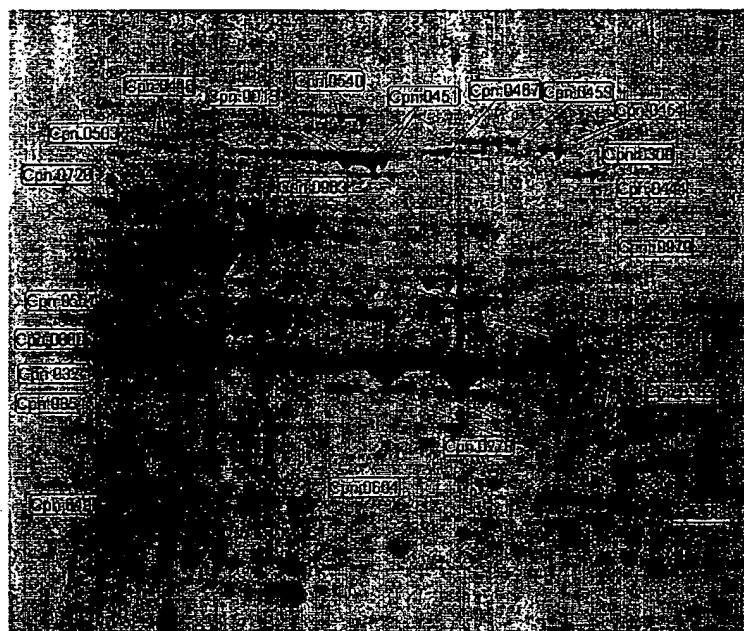
(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European

(71) Applicant (for all designated States except US): CHIRON S.P.A. [IT/IT]; Via Fiorentina, 1, I-53100 Siena (IT).

[Continued on next page]

(54) Title: IMMUNISATION AGAINST *CHLAMYDIA PNEUMONIAE*



(57) Abstract: The published genomic of *Chlamydia pneumoniae* reveals over 1000 putative encoded proteins but does not itself indicate which of these might be useful antigens for immunisation and vaccination or for diagnosis. This difficulty is addressed by the invention, which provides a number of *C. pneumoniae* protein sequences suitable for vaccine production and development and/or for diagnostic purposes.

WO 02/02606 A2



patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

*For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.*

**Published:**

- *without international search report and to be republished upon receipt of that report*

## IMMUNISATION AGAINST *CHLAMYDIA PNEUMONIAE*

All documents cited herein are incorporated by reference in their entirety.

### TECHNICAL FIELD

This invention is in the field of immunisation against chlamydial infection, in particular against infection by *Chlamydia pneumoniae*.

### BACKGROUND ART

*Chlamydiae* are obligate intracellular parasites of eukaryotic cells which are responsible for endemic sexually transmitted infections and various other disease syndromes. They occupy an exclusive eubacterial phylogenetic branch, having no close relationship to any other known organisms – they are classified in their own order (*Chlamydiales*) which contains a single family (*Chlamydiaceae*) which in turn contains a single genus (*Chlamydia*). A particular characteristic of the *Chlamydiae* is their unique life cycle, in which the bacterium alternates between two morphologically distinct forms: an extracellular infective form (elementary bodies, EB) and an intracellular non-infective form (reticulate bodies, RB). The life cycle is completed with the re-organization of RB into EB, which subsequently leave the disrupted host cell ready to infect further cells.

Four chlamydial species are currently known – *C.trachomatis*, *C.pneumoniae*, *C.pecorum* and *C.psittaci* [e.g. Raulston (1995) *Mol Microbiol* 15:607-616; Everett (2000) *Vet Microbiol* 75:109-126]. *C.pneumoniae* is closely related to *C.trachomatis*, as the whole genome comparison of at least two isolates from each species has shown [Kalman *et al.* (1999) *Nature Genetics* 21:385-389; Read *et al.* (2000) *Nucleic Acids Res* 28:1397-406; Stephens *et al.* (1998) *Science* 282:754-759]. Based on surface reaction with patient immune sera, the current view is that only one serotype of *C.pneumoniae* exists world-wide.

*C.pneumoniae* is a common cause of human respiratory disease. It was first isolated from the conjunctiva of a child in Taiwan in 1965, and was established as a major respiratory pathogen in 1983. In the USA, *C.pneumoniae* causes approximately 10% of community-acquired pneumonia and 5% of pharyngitis, bronchitis, and sinusitis.

More recently, the spectrum of *C.pneumoniae* infections has been extended to include atherosclerosis, coronary heart disease, carotid artery stenosis, myocardial infarction, cerebrovascular disease, aortic aneurysm, claudication, and stroke. The association of *C.pneumoniae* with atherosclerosis is corroborated by the presence of the organism in atherosclerotic lesions throughout the arterial tree and the near absence of the organism in healthy arterial tissue. *C.pneumoniae* has also been isolated from coronary and carotid atheromatous plaques. The bacterium has also been associated with other acute and chronic respiratory diseases (e.g. otitis media, chronic obstructive pulmonary disease, pulmonary exacerbation of cystic fibrosis) as a result of sero-epidemiologic observations, case reports, isolation or direct detection of the organism in specimens, and successful



response to anti-chlamydial antibiotics. To determine whether chronic infection plays a role in initiation or progression of disease, intervention studies in humans have been initiated, and animal models of *C.pneumoniae* infection have been developed.

Considerable knowledge of the epidemiology of *C.pneumoniae* infection has been derived from serologic studies using the *C.pneumoniae*-specific microimmunofluorescence test. Infection is ubiquitous, and it is estimated that virtually everyone is infected at some point in life, with common re-infection. Antibodies against *C.pneumoniae* are rare in children under the age of 5, except in developing and tropical countries. Antibody prevalence increases rapidly at ages 5 to 14, reaching 50% at the age of 20, and continuing to increase slowly to ~80% by age 70.

A current hypothesis is that *C.pneumoniae* can persist in an asymptomatic low-grade infection in very large sections of the human population. When this condition occurs, it is believed that the presence of *C.pneumoniae*, and/or the effects of the host reaction to the bacterium, can cause or help progress of cardiovascular illness.

It is not yet clear whether *C.pneumoniae* is actually a causative agent of cardiovascular disease, or whether it is just artefactually associated with it. It has been shown, however, that *C.pneumoniae* infection can induce LDL oxidation by human monocytes [Kalayoglu *et al.* (1999) *J. Infect. Dis.* 180:780-90; Kalayoglu *et al.* (1999) *Am. Heart J.* 138:S488-490]. As LDL oxidation products are highly atherogenic, this observation provides a possible mechanism whereby *C.pneumoniae* may cause atheromatous degeneration. If a causative effect is confirmed, vaccination (prophylactic and therapeutic) will be universally recommended.

Genomic sequence information has been published for *C.pneumoniae* [Kalman *et al.* (1999) *supra*; Read *et al.* (2000) *supra*; Shirai *et al.* (2000) *J. Infect. Dis.* 181(Suppl 3):S524-S527; WO99/27105; WO00/27994] and is available from GenBank. Sequencing efforts have not, however, focused on vaccination, and the availability of genomic sequence does not in itself indicate which of the >1000 genes might encode useful antigens for immunisation and vaccination. WO99/27105, for instance, implies that every one of the 1296 ORFs identified in the *C.pneumoniae* strain CM1 genome is a useful vaccine antigen.

It is thus an object of the present invention to identify antigens useful for vaccine production and development from amongst the many proteins present in *C.pneumoniae*. It is a further object to identify antigens useful for diagnosis (e.g. immunodiagnosis) of *C.pneumoniae*.

## DISCLOSURE OF THE INVENTION

The invention provides proteins comprising the *C.pneumoniae* amino acid sequences disclosed in the examples.

It also provides proteins comprising sequences which share at least x% sequence identity with the *C.pneumoniae* amino acid sequences disclosed in the examples. Depending on the particular

sequence,  $x$  is preferably 50% or more (e.g. 60%, 70%, 80%, 90%, 95%, 99% or more). These include mutants and allelic variants. Typically, 50% identity or more between two proteins is considered to be an indication of functional equivalence. Identity between proteins is preferably determined by the Smith-Waterman homology search algorithm as implemented in the MPSRCH  
5 program (Oxford Molecular), using an affine gap search with parameters *gap open penalty*=12 and *gap extension penalty*=1.

The invention further provides proteins comprising fragments of the *C.pneumoniae* amino acid sequences disclosed in the examples. The fragments should comprise at least  $n$  consecutive amino acids from the sequences and, depending on the particular sequence,  $n$  is 7 or more (e.g. 8, 10, 12,  
10 14, 16, 18, 20, 30, 40, 50, 75, 100 or more). Preferably the fragments comprise one or more epitope(s) from the sequence. Other preferred fragments omit a signal peptide.

The proteins of the invention can, of course, be prepared by various means (e.g. native expression, recombinant expression, purification from cell culture, chemical synthesis etc.) and in various forms (e.g. native, fusions etc.). They are preferably prepared in substantially pure form (ie. substantially  
15 free from other *C.pneumoniae* or host cell proteins). Heterologous expression in *E.coli* is a preferred preparative route.

According to a further aspect, the invention provides nucleic acid comprising the *C.pneumoniae* nucleotide sequences disclosed in the examples. In addition, the invention provides nucleic acid comprising sequences which share at least  $x\%$  sequence identity with the *C.pneumoniae* nucleotide  
20 sequences disclosed in the examples. Depending on the particular sequence,  $x$  is preferably 50% or more (e.g. 60%, 70%, 80%, 90%, 95%, 99% or more).

Furthermore, the invention provides nucleic acid which can hybridise to the *C.pneumoniae* nucleic acid disclosed in the examples, preferably under "high stringency" conditions (e.g. 65°C in a 0.1xSSC, 0.5% SDS solution).

25 Nucleic acid comprising fragments of these sequences are also provided. These should comprise at least  $n$  consecutive nucleotides from the *C.pneumoniae* sequences and, depending on the particular sequence,  $n$  is 10 or more (e.g. 12, 14, 15, 18, 20, 25, 30, 35, 40, 50, 75, 100, 200, 300 or more).

According to a further aspect, the invention provides nucleic acid encoding the proteins and protein fragments of the invention.

30 It should also be appreciated that the invention provides nucleic acid comprising sequences complementary to those described above (e.g. for antisense or probing purposes).

Nucleic acid according to the invention can, of course, be prepared in many ways (e.g. by chemical synthesis, from genomic or cDNA libraries, from the organism itself etc.) and can take various forms (e.g. single stranded, double stranded, vectors, probes etc.).

In addition, the term "nucleic acid" includes DNA and RNA, and also their analogues, such as those containing modified backbones, and also peptide nucleic acids (PNA) *etc.*

According to a further aspect, the invention provides vectors comprising nucleotide sequences of the invention (*e.g.* cloning or expression vectors) and host cells transformed therewith.

- 5 According to a further aspect, the invention provides immunogenic compositions comprising protein and/or nucleic acid according to the invention. These compositions are suitable for immunisation and vaccination purposes. Vaccines of the invention may be prophylactic or therapeutic, and will typically comprise an antigen which can induce antibodies capable of inhibiting (a) chlamydial adhesion, (b) chlamydial entry, and/or (c) successful replication within the host cell. The vaccines  
10 preferably induce any cell-mediated T-cell responses which are necessary for chlamydial clearance from the host.

- The invention also provides nucleic acid or protein according to the invention for use as medicaments (*e.g.* as vaccines). It also provides the use of nucleic acid or protein according to the invention in the manufacture of a medicament (*e.g.* a vaccine or an immunogenic composition) for  
15 treating or preventing infection due to *C.pneumoniae*.

The invention also provides a method of treating (*e.g.* immunising) a patient, comprising administering to the patient a therapeutically effective amount of nucleic acid or protein according to the invention.

According to further aspects, the invention provides various processes.

- 20 A process for producing proteins of the invention is provided, comprising the step of culturing a host cell according to the invention under conditions which induce protein expression.

A process for producing protein or nucleic acid of the invention is provided, wherein the protein or nucleic acid is synthesised in part or in whole using chemical means.

- A process for detecting *C.pneumoniae* in a sample is provided, wherein the sample is contacted with  
25 an antibody which binds to a protein of the invention.

A summary of standard techniques and procedures which may be employed in order to perform the invention (*e.g.* to utilise the disclosed sequences for immunisation) follows. This summary is not a limitation on the invention but, rather, gives examples that may be used, but are not required.

#### General

- 30 The practice of the present invention will employ, unless otherwise indicated, conventional techniques of molecular biology, microbiology, recombinant DNA, and immunology, which are within the skill of the art. Such techniques are explained fully in the literature *e.g.* Sambrook *Molecular Cloning; A Laboratory Manual, Second Edition* (1989) and *Third Edition* (2001); *DNA Cloning, Volumes I and ii* (D.N Glover ed. 1985); *Oligonucleotide Synthesis* (M.J. Gait ed, 1984); *Nucleic Acid Hybridization* (B.D. Hames & S.J. Higgins eds.  
35 1984); *Transcription and Translation* (B.D. Hames & S.J. Higgins eds. 1984); *Animal Cell Culture* (R.I.

- 5 Freshney ed. 1986); *Immobilized Cells and Enzymes* (IRL Press, 1986); B. Perbal, *A Practical Guide to Molecular Cloning* (1984); the *Methods in Enzymology* series (Academic Press, Inc.), especially volumes 154 & 155; *Gene Transfer Vectors for Mammalian Cells* (J.H. Miller and M.P. Calos eds. 1987, Cold Spring Harbor Laboratory); Mayer and Walker, eds. (1987), *Immunochemical Methods in Cell and Molecular Biology* (Academic Press, London); Scopes, (1987) *Protein Purification: Principles and Practice*, Second Edition (Springer-Verlag, N.Y.), and *Handbook of Experimental Immunology, Volumes I-IV* (D.M. Weir and C. C. Blackwell eds 1986).

Standard abbreviations for nucleotides and amino acids are used in this specification.

#### Definitions

- 10 A composition containing X is "substantially free of" Y when at least 85% by weight of the total X+Y in the composition is X. Preferably, X comprises at least about 90% by weight of the total of X+Y in the composition, more preferably at least about 95% or even 99% by weight.

The term "comprising" means "including" as well as "consisting" e.g. a composition "comprising" X may consist exclusively of X or may include something additional to X, such as X+Y.

- 15 The term "heterologous" refers to two biological components that are not found together in nature. The components may be host cells, genes, or regulatory regions, such as promoters. Although the heterologous components are not found together in nature, they can function together, as when a promoter heterologous to a gene is operably linked to the gene. Another example is where a Chlamydial sequence is heterologous to a mouse host cell. A further examples would be two epitopes from the same or different proteins which have been  
20 assembled in a single protein in an arrangement not found in nature.

- An "origin of replication" is a polynucleotide sequence that initiates and regulates replication of polynucleotides, such as an expression vector. The origin of replication behaves as an autonomous unit of polynucleotide replication within a cell, capable of replication under its own control. An origin of replication may be needed for a vector to replicate in a particular host cell. With certain origins of replication, an expression vector can be  
25 reproduced at a high copy number in the presence of the appropriate proteins within the cell. Examples of origins are the autonomously replicating sequences, which are effective in yeast; and the viral T-antigen, effective in COS-7 cells.

- A "mutant" sequence is defined as DNA, RNA or amino acid sequence differing from but having sequence identity with the native or disclosed sequence. Depending on the particular sequence, the degree of sequence  
30 identity between the native or disclosed sequence and the mutant sequence is preferably greater than 50% (e.g. 60%, 70%, 80%, 90%, 95%, 99% or more, calculated using the Smith-Waterman algorithm as described above). As used herein, an "allelic variant" of a nucleic acid molecule, or region, for which nucleic acid sequence is provided herein is a nucleic acid molecule, or region, that occurs essentially at the same locus in the genome of another or second isolate, and that, due to natural variation caused by, for example, mutation or recombination,  
35 has a similar but not identical nucleic acid sequence. A coding region allelic variant typically encodes a protein having similar activity to that of the protein encoded by the gene to which it is being compared. An allelic variant can also comprise an alteration in the 5' or 3' untranslated regions of the gene, such as in regulatory control regions (e.g. see US patent 5,753,235).

Expression systems

The Chlamydial nucleotide sequences can be expressed in a variety of different expression systems; for example those used with mammalian cells, baculoviruses, plants, bacteria, and yeast.

i. Mammalian Systems

- 5 Mammalian expression systems are known in the art. A mammalian promoter is any DNA sequence capable of binding mammalian RNA polymerase and initiating the downstream (3') transcription of a coding sequence (e.g. structural gene) into mRNA. A promoter will have a transcription initiating region, which is usually placed proximal to the 5' end of the coding sequence, and a TATA box, usually located 25-30 base pairs (bp) upstream of the transcription initiation site. The TATA box is thought to direct RNA polymerase II to begin RNA  
10 synthesis at the correct site. A mammalian promoter will also contain an upstream promoter element, usually located within 100 to 200 bp upstream of the TATA box. An upstream promoter element determines the rate at which transcription is initiated and can act in either orientation [Sambrook et al. (1989) "Expression of Cloned Genes in Mammalian Cells." In *Molecular Cloning: A Laboratory Manual*, 2nd ed.].

- 15 Mammalian viral genes are often highly expressed and have a broad host range; therefore sequences encoding mammalian viral genes provide particularly useful promoter sequences. Examples include the SV40 early promoter, mouse mammary tumor virus LTR promoter, adenovirus major late promoter (Ad MLP), and herpes simplex virus promoter. In addition, sequences derived from non-viral genes, such as the murine metallothionein gene, also provide useful promoter sequences. Expression may be either constitutive or regulated (inducible), depending on the promoter can be induced with glucocorticoid in hormone-responsive  
20 cells.

- The presence of an enhancer element (enhancer), combined with the promoter elements described above, will usually increase expression levels. An enhancer is a regulatory DNA sequence that can stimulate transcription up to 1000-fold when linked to homologous or heterologous promoters, with synthesis beginning at the normal RNA start site. Enhancers are also active when they are placed upstream or downstream from the transcription  
25 initiation site, in either normal or flipped orientation, or at a distance of more than 1000 nucleotides from the promoter [Maniatis et al. (1987) *Science* 236:1237; Alberts et al. (1989) *Molecular Biology of the Cell*, 2nd ed.]. Enhancer elements derived from viruses may be particularly useful, because they usually have a broader host range. Examples include the SV40 early gene enhancer [Dijkema et al (1985) *EMBO J.* 4:761] and the enhancer/promoters derived from the long terminal repeat (LTR) of the Rous Sarcoma Virus [Gorman et al.  
30 (1982) *PNAS USA* 79:6777] and from human cytomegalovirus [Boshart et al. (1985) *Cell* 41:521]. Additionally, some enhancers are regulatable and become active only in the presence of an inducer, such as a hormone or metal ion [Sassone-Corsi and Borelli (1986) *Trends Genet.* 2:215; Maniatis et al. (1987) *Science* 236:1237].

- A DNA molecule may be expressed intracellularly in mammalian cells. A promoter sequence may be directly linked with the DNA molecule, in which case the first amino acid at the N-terminus of the recombinant protein  
35 will always be a methionine, which is encoded by the ATG start codon. If desired, the N-terminus may be cleaved from the protein by *in vitro* incubation with cyanogen bromide.

Alternatively, foreign proteins can also be secreted from the cell into the growth media by creating chimeric DNA molecules that encode a fusion protein comprised of a leader sequence fragment that provides for secretion of the foreign protein in mammalian cells. Preferably, there are processing sites encoded between the leader

fragment and the foreign gene that can be cleaved either *in vivo* or *in vitro*. The leader sequence fragment usually encodes a signal peptide comprised of hydrophobic amino acids which direct the secretion of the protein from the cell. The adenovirus tripartite leader is an example of a leader sequence that provides for secretion of a foreign protein in mammalian cells.

- 5 Usually, transcription termination and polyadenylation sequences recognized by mammalian cells are regulatory regions located 3' to the translation stop codon and thus, together with the promoter elements, flank the coding sequence. The 3' terminus of the mature mRNA is formed by site-specific post-transcriptional cleavage and polyadenylation [Birnstiel et al. (1985) *Cell* 41:349; Proudfoot and Whitelaw (1988) "Termination and 3' end processing of eukaryotic RNA. In *Transcription and splicing* (ed. B.D. Hames and D.M. Glover); Proudfoot
- 10 (1989) *Trends Biochem. Sci.* 14:105]. These sequences direct the transcription of an mRNA which can be translated into the polypeptide encoded by the DNA. Examples of transcription terminator/polyadenylation signals include those derived from SV40 [Sambrook et al (1989) "Expression of cloned genes in cultured mammalian cells." In *Molecular Cloning: A Laboratory Manual*].

- Usually, the above described components, comprising a promoter, polyadenylation signal, and transcription termination sequence are put together into expression constructs. Enhancers, introns with functional splice donor and acceptor sites, and leader sequences may also be included in an expression construct, if desired. Expression constructs are often maintained in a replicon, such as an extrachromosomal element (e.g. plasmids) capable of stable maintenance in a host, such as mammalian cells or bacteria. Mammalian replication systems include those derived from animal viruses, which require trans-acting factors to replicate. For example, plasmids containing
- 15 the replication systems of papovaviruses, such as SV40 [Gluzman (1981) *Cell* 23:175] or polyomavirus, replicate to extremely high copy number in the presence of the appropriate viral T antigen. Additional examples of mammalian replicons include those derived from bovine papillomavirus and Epstein-Barr virus. Additionally, the replicon may have two replicaton systems, thus allowing it to be maintained, for example, in mammalian cells for expression and in a prokaryotic host for cloning and amplification. Examples of such mammalian-
- 20 bacteria shuttle vectors include pMT2 [Kaufman et al. (1989) *Mol. Cell. Biol.* 9:946] and pHEBO [Shimizu et al. (1986) *Mol. Cell. Biol.* 6:1074].

- The transformation procedure used depends upon the host to be transformed. Methods for introduction of heterologous polynucleotides into mammalian cells are known in the art and include dextran-mediated transfection, calcium phosphate precipitation, polybrene-mediated transfection, protoplast fusion,
- 25 electroporation, encapsulation of polynucleotide(s) in liposomes, direct microinjection of the DNA into nuclei.

Mammalian cell lines available as hosts for expression are known in the art and include many immortalized cell lines available from the American Type Culture Collection (ATCC), including but not limited to, Chinese hamster ovary (CHO) cells, HeLa cells, baby hamster kidney (BHK) cells, monkey kidney cells (COS), human hepatocellular carcinoma cells (e.g. Hep G2), and a number of other cell lines.

## 35 ii. Baculovirus Systems

- The polynucleotide encoding the protein can also be inserted into a suitable insect expression vector, and is operably linked to the control elements within that vector. Vector construction employs techniques which are known in the art. Generally, the components of the expression system include a transfer vector, usually a bacterial plasmid, which contains both a fragment of the baculovirus genome, and a convenient restriction site
- 40 for insertion of the heterologous gene or genes to be expressed; a wild type baculovirus with a sequence

homologous to the baculovirus-specific fragment in the transfer vector (this allows for the homologous recombination of the heterologous gene in to the baculovirus genome); and appropriate insect host cells and growth media.

After inserting the DNA sequence encoding the protein into the transfer vector, the vector and the wild type viral genome are transfected into an insect host cell where the vector and viral genome are allowed to recombine. The packaged recombinant virus is expressed and recombinant plaques are identified and purified. Materials and methods for baculovirus/insect cell expression systems are commercially available in kit form from, *inter alia*, Invitrogen, San Diego CA ("MaxBac" kit). These techniques are generally known to those skilled in the art and fully described in Summers and Smith, *Texas Agricultural Experiment Station Bulletin No. 1555* (1987) (hereinafter "Summers and Smith").

Prior to inserting the DNA sequence encoding the protein into the baculovirus genome, the above described components, comprising a promoter, leader (if desired), coding sequence of interest, and transcription termination sequence, are usually assembled into an intermediate transplacement construct (transfer vector). This construct may contain a single gene and operably linked regulatory elements; multiple genes, each with its own set of operably linked regulatory elements; or multiple genes, regulated by the same set of regulatory elements. Intermediate transplacement constructs are often maintained in a replicon, such as an extrachromosomal element (e.g. plasmids) capable of stable maintenance in a host, such as a bacterium. The replicon will have a replication system, thus allowing it to be maintained in a suitable host for cloning and amplification.

Currently, the most commonly used transfer vector for introducing foreign genes into AcNPV is pAc373. Many other vectors, known to those of skill in the art, have also been designed. These include, for example, pVL985 (which alters the polyhedrin start codon from ATG to ATT, and which introduces a BamHI cloning site 32 basepairs downstream from the ATT; see Luckow and Summers, *Virology* (1989) 17:31.

The plasmid usually also contains the polyhedrin polyadenylation signal (Miller et al. (1988) *Ann. Rev. Microbiol.*, 42:177) and a prokaryotic ampicillin-resistance (*amp*) gene and origin of replication for selection and propagation in *E. coli*.

Baculovirus transfer vectors usually contain a baculovirus promoter. A baculovirus promoter is any DNA sequence capable of binding a baculovirus RNA polymerase and initiating the downstream (5' to 3') transcription of a coding sequence (e.g. structural gene) into mRNA. A promoter will have a transcription initiation region which is usually placed proximal to the 5' end of the coding sequence. This transcription initiation region usually includes an RNA polymerase binding site and a transcription initiation site. A baculovirus transfer vector may also have a second domain called an enhancer, which, if present, is usually distal to the structural gene. Expression may be either regulated or constitutive.

Structural genes, abundantly transcribed at late times in a viral infection cycle, provide particularly useful promoter sequences. Examples include sequences derived from the gene encoding the viral polyhedron protein, Friesen et al., (1986) "The Regulation of Baculovirus Gene Expression," in: *The Molecular Biology of Baculoviruses* (ed. Walter Doerfler); EPO Publ. Nos. 127 839 and 155 476; and the gene encoding the p10 protein, Vlak et al., (1988), *J. Gen. Virol.* 69:765.

DNA encoding suitable signal sequences can be derived from genes for secreted insect or baculovirus proteins, such as the baculovirus polyhedrin gene (Carbonell et al. (1988) *Gene*, 73:409). Alternatively, since the signals

for mammalian cell posttranslational modifications (such as signal peptide cleavage, proteolytic cleavage, and phosphorylation) appear to be recognized by insect cells, and the signals required for secretion and nuclear accumulation also appear to be conserved between the invertebrate cells and vertebrate cells, leaders of non-insect origin, such as those derived from genes encoding human  $\alpha$ -interferon, Maeda et al., (1985), *Nature* 315:592; human gastrin-releasing peptide, Lebacqz-Verheyden et al., (1988), *Molec. Cell. Biol.* 8:3129; human IL-2, Smith et al., (1985) *Proc. Nat'l Acad. Sci. USA*, 82:8404; mouse IL-3, (Miyajima et al., (1987) *Gene* 58:273; and human glucocerebrosidase, Martin et al. (1988) *DNA*, 7:99, can also be used to provide for secretion in insects.

A recombinant polypeptide or polyprotein may be expressed intracellularly or, if it is expressed with the proper regulatory sequences, it can be secreted. Good intracellular expression of nonfused foreign proteins usually requires heterologous genes that ideally have a short leader sequence containing suitable translation initiation signals preceding an ATG start signal. If desired, methionine at the N-terminus may be cleaved from the mature protein by *in vitro* incubation with cyanogen bromide.

Alternatively, recombinant polyproteins or proteins which are not naturally secreted can be secreted from the insect cell by creating chimeric DNA molecules that encode a fusion protein comprised of a leader sequence fragment that provides for secretion of the foreign protein in insects. The leader sequence fragment usually encodes a signal peptide comprised of hydrophobic amino acids which direct the translocation of the protein into the endoplasmic reticulum.

After insertion of the DNA sequence and/or the gene encoding the expression product precursor of the protein, an insect cell host is co-transformed with the heterologous DNA of the transfer vector and the genomic DNA of wild type baculovirus -- usually by co-transfection. The promoter and transcription termination sequence of the construct will usually comprise a 2-5kb section of the baculovirus genome. Methods for introducing heterologous DNA into the desired site in the baculovirus virus are known in the art. (See Summers and Smith *supra*; Ju et al. (1987); Smith et al., *Mol. Cell. Biol.* (1983) 3:2156; and Luckow and Summers (1989)). For example, the insertion can be into a gene such as the polyhedrin gene, by homologous double crossover recombination; insertion can also be into a restriction enzyme site engineered into the desired baculovirus gene. Miller et al., (1989), *Bioessays* 4:91. The DNA sequence, when cloned in place of the polyhedrin gene in the expression vector, is flanked both 5' and 3' by polyhedrin-specific sequences and is positioned downstream of the polyhedrin promoter.

The newly formed baculovirus expression vector is subsequently packaged into an infectious recombinant baculovirus. Homologous recombination occurs at low frequency (between ~1% and ~5%); thus, the majority of the virus produced after cotransfection is still wild-type virus. Therefore, a method is necessary to identify recombinant viruses. An advantage of the expression system is a visual screen allowing recombinant viruses to be distinguished. The polyhedrin protein, which is produced by the native virus, is produced at very high levels in the nuclei of infected cells at late times after viral infection. Accumulated polyhedrin protein forms occlusion bodies that also contain embedded particles. These occlusion bodies, up to 15 $\mu$ m in size, are highly refractile, giving them a bright shiny appearance that is readily visualized under the light microscope. Cells infected with recombinant viruses lack occlusion bodies. To distinguish recombinant virus from wild-type virus, the transfection supernatant is plaqued onto a monolayer of insect cells by techniques known to those skilled in the art. Namely, the plaques are screened under the light microscope for the presence (indicative of wild-type virus)



or absence (indicative of recombinant virus) of occlusion bodies. "Current Protocols in Microbiology" Vol. 2 (Ausubel et al. eds) at 16.8 (Supp. 10, 1990); Summers & Smith, *supra*; Miller et al. (1989).

Recombinant baculovirus expression vectors have been developed for infection into several insect cells. For example, recombinant baculoviruses have been developed for, *inter alia*: *Aedes aegypti*, *Autographa californica*, *Bombyx mori*, *Drosophila melanogaster*, *Spodoptera frugiperda*, and *Trichoplusia ni* (WO 89/046699; Carbonell et al., (1985) *J. Virol.* 56:153; Wright (1986) *Nature* 321:718; Smith et al., (1983) *Mol. Cell. Biol.* 3:2156; and see generally, Fraser, et al. (1989) *In Vitro Cell. Dev. Biol.* 25:225).

Cells and cell culture media are commercially available for both direct and fusion expression of heterologous polypeptides in a baculovirus/expression system; cell culture technology is generally known to those skilled in the art. See, e.g. Summers and Smith *supra*.

The modified insect cells may then be grown in an appropriate nutrient medium, which allows for stable maintenance of the plasmid(s) present in the modified insect host. Where the expression product gene is under inducible control, the host may be grown to high density, and expression induced. Alternatively, where expression is constitutive, the product will be continuously expressed into the medium and the nutrient medium must be continuously circulated, while removing the product of interest and augmenting depleted nutrients. The product may be purified by such techniques as chromatography, e.g. HPLC, affinity chromatography, ion exchange chromatography, etc.; electrophoresis; density gradient centrifugation; solvent extraction, or the like. As appropriate, the product may be further purified, as required, so as to remove substantially any insect proteins which are also secreted in the medium or result from lysis of insect cells, so as to provide a product which is at least substantially free of host debris, e.g. proteins, lipids and polysaccharides.

In order to obtain protein expression, recombinant host cells derived from the transformants are incubated under conditions which allow expression of the recombinant protein encoding sequence. These conditions will vary, dependent upon the host cell selected. However, the conditions are readily ascertainable to those of ordinary skill in the art, based upon what is known in the art.

### iii. Plant Systems

There are many plant cell culture and whole plant genetic expression systems known in the art. Exemplary plant cellular genetic expression systems include those described in patents, such as: US 5,693,506; US 5,659,122; and US 5,608,143. Additional examples of genetic expression in plant cell culture has been described by Zenk, *Phytochemistry* 30:3861-3863 (1991). Descriptions of plant protein signal peptides may be found in addition to the references described above in Vaulcombe et al., *Mol. Gen. Genet.* 209:33-40 (1987); Chandler et al., *Plant Molecular Biology* 3:407-418 (1984); Rogers, *J. Biol. Chem.* 260:3731-3738 (1985); Rothstein et al., *Gene* 55:353-356 (1987); Whittier et al., *Nucleic Acids Research* 15:2515-2535 (1987); Wirsal et al., *Molecular Microbiology* 3:3-14 (1989); Yu et al., *Gene* 122:247-253 (1992). A description of the regulation of plant gene expression by the phytohormone, gibberellic acid and secreted enzymes induced by gibberellic acid can be found in R.L. Jones and J. MacMillan, *Gibberellins*: in: *Advanced Plant Physiology*, Malcolm B. Wilkins, ed., 1984 Pitman Publishing Limited, London, pp. 21-52. References that describe other metabolically-regulated genes: Sheen, *Plant Cell*, 2:1027-1038(1990); Maas et al., *EMBO J.* 9:3447-3452 (1990); Benkel and Hickey, *Proc. Natl. Acad. Sci.* 84:1337-1339 (1987)

Typically, using techniques known in the art, a desired polynucleotide sequence is inserted into an expression cassette comprising genetic regulatory elements designed for operation in plants. The expression cassette is inserted into a desired expression vector with companion sequences upstream and downstream from the expression cassette suitable for expression in a plant host. The companion sequences will be of plasmid or viral origin and provide necessary characteristics to the vector to permit the vectors to move DNA from an original cloning host, such as bacteria, to the desired plant host. The basic bacterial/plant vector construct will preferably provide a broad host range prokaryote replication origin; a prokaryote selectable marker; and, for *Agrobacterium* transformations, T DNA sequences for *Agrobacterium*-mediated transfer to plant chromosomes. Where the heterologous gene is not readily amenable to detection, the construct will preferably also have a selectable marker gene suitable for determining if a plant cell has been transformed. A general review of suitable markers, for example for the members of the grass family, is found in Wilmink and Dons, 1993, *Plant Mol. Biol. Reprtr*, 11(2):165-185.

Sequences suitable for permitting integration of the heterologous sequence into the plant genome are also recommended. These might include transposon sequences and the like for homologous recombination as well as Ti sequences which permit random insertion of a heterologous expression cassette into a plant genome. Suitable prokaryote selectable markers include resistance toward antibiotics such as ampicillin or tetracycline. Other DNA sequences encoding additional functions may also be present in the vector, as is known in the art.

The nucleic acid molecules of the subject invention may be included into an expression cassette for expression of the protein(s) of interest. Usually, there will be only one expression cassette, although two or more are feasible. The recombinant expression cassette will contain in addition to the heterologous protein encoding sequence the following elements, a promoter region, plant 5' untranslated sequences, initiation codon depending upon whether or not the structural gene comes equipped with one, and a transcription and translation termination sequence. Unique restriction enzyme sites at the 5' and 3' ends of the cassette allow for easy insertion into a pre-existing vector.

A heterologous coding sequence may be for any protein relating to the present invention. The sequence encoding the protein of interest will encode a signal peptide which allows processing and translocation of the protein, as appropriate, and will usually lack any sequence which might result in the binding of the desired protein of the invention to a membrane. Since, for the most part, the transcriptional initiation region will be for a gene which is expressed and translocated during germination, by employing the signal peptide which provides for translocation, one may also provide for translocation of the protein of interest. In this way, the protein(s) of interest will be translocated from the cells in which they are expressed and may be efficiently harvested. Typically secretion in seeds are across the aleurone or scutellar epithelium layer into the endosperm of the seed. While it is not required that the protein be secreted from the cells in which the protein is produced, this facilitates the isolation and purification of the recombinant protein.

Since the ultimate expression of the desired gene product will be in a eucaryotic cell it is desirable to determine whether any portion of the cloned gene contains sequences which will be processed out as introns by the host's splicosome machinery. If so, site-directed mutagenesis of the "intron" region may be conducted to prevent losing a portion of the genetic message as a false intron code, Reed and Maniatis, *Cell* 41:95-105, 1985.

The vector can be microinjected directly into plant cells by use of micropipettes to mechanically transfer the recombinant DNA. Crossway, *Mol. Gen. Genet*, 202:179-185, 1985. The genetic material may also be

transferred into the plant cell by using polyethylene glycol, Krens, et al., *Nature*, 296, 72-74, 1982. Another method of introduction of nucleic acid segments is high velocity ballistic penetration by small particles with the nucleic acid either within the matrix of small beads or particles, or on the surface, Klein, et al., *Nature*, 327, 70-73, 1987 and Knudsen and Muller, 1991, *Planta*, 185:330-336 teaching particle bombardment of barley endosperm to create transgenic barley. Yet another method of introduction would be fusion of protoplasts with other entities, either minicells, cells, lysosomes or other fusible lipid-surfaced bodies, Fraley, et al., *Proc. Natl. Acad. Sci. USA*, 79, 1859-1863, 1982.

The vector may also be introduced into the plant cells by electroporation. (Fromm et al., *Proc. Natl. Acad. Sci. USA* 82:5824, 1985). In this technique, plant protoplasts are electroporated in the presence of plasmids containing the gene construct. Electrical impulses of high field strength reversibly permeabilize biomembranes allowing the introduction of the plasmids. Electroporated plant protoplasts reform the cell wall, divide, and form plant callus.

All plants from which protoplasts can be isolated and cultured to give whole regenerated plants can be transformed by the present invention so that whole plants are recovered which contain the transferred gene. It is known that practically all plants can be regenerated from cultured cells or tissues, including but not limited to all major species of sugarcane, sugar beet, cotton, fruit and other trees, legumes and vegetables. Some suitable plants include, for example, species from the genera *Fragaria*, *Lotus*, *Medicago*, *Onobrychis*, *Trifolium*, *Trigonella*, *Vigna*, *Citrus*, *Linum*, *Geranium*, *Manihot*, *Daucus*, *Arabidopsis*, *Brassica*, *Raphanus*, *Sinapis*, *Atropa*, *Capsicum*, *Datura*, *Hyoscyamus*, *Lycopersion*, *Nicotiana*, *Solanum*, *Petunia*, *Digitalis*, *Majorana*, *Cichorium*, *Helianthus*, *Lactuca*, *Bromus*, *Asparagus*, *Antirrhinum*, *Hieracallis*, *Nemesia*, *Pelargonium*, *Panicum*, *Pennisetum*, *Ranunculus*, *Senecio*, *Salpiglossis*, *Cucumis*, *Browaalia*, *Glycine*, *Lolium*, *Zea*, *Triticum*, *Sorghum*, and *Datura*.

Means for regeneration vary from species to species of plants, but generally a suspension of transformed protoplasts containing copies of the heterologous gene is first provided. Callus tissue is formed and shoots may be induced from callus and subsequently rooted. Alternatively, embryo formation can be induced from the protoplast suspension. These embryos germinate as natural embryos to form plants. The culture media will generally contain various amino acids and hormones, such as auxin and cytokinins. It is also advantageous to add glutamic acid and proline to the medium, especially for such species as corn and alfalfa. Shoots and roots normally develop simultaneously. Efficient regeneration will depend on the medium, on the genotype, and on the history of the culture. If these three variables are controlled, then regeneration is fully reproducible and repeatable.

In some plant cell culture systems, the desired protein of the invention may be excreted or alternatively, the protein may be extracted from the whole plant. Where the desired protein of the invention is secreted into the medium, it may be collected. Alternatively, the embryos and embryoless-half seeds or other plant tissue may be mechanically disrupted to release any secreted protein between cells and tissues. The mixture may be suspended in a buffer solution to retrieve soluble proteins. Conventional protein isolation and purification methods will be then used to purify the recombinant protein. Parameters of time, temperature pH, oxygen, and volumes will be adjusted through routine methods to optimize expression and recovery of heterologous protein.

#### iv. Bacterial Systems

Bacterial expression techniques are known in the art. A bacterial promoter is any DNA sequence capable of binding bacterial RNA polymerase and initiating the downstream (3') transcription of a coding sequence (e.g. structural gene) into mRNA. A promoter will have a transcription initiation region which is usually placed proximal to the 5' end of the coding sequence. This transcription initiation region usually includes an RNA polymerase binding site and a transcription initiation site. A bacterial promoter may also have a second domain called an operator, that may overlap an adjacent RNA polymerase binding site at which RNA synthesis begins. The operator permits negative regulated (inducible) transcription, as a gene repressor protein may bind the operator and thereby inhibit transcription of a specific gene. Constitutive expression may occur in the absence of negative regulatory elements, such as the operator. In addition, positive regulation may be achieved by a gene activator protein binding sequence, which, if present is usually proximal (5') to the RNA polymerase binding sequence. An example of a gene activator protein is the catabolite activator protein (CAP), which helps initiate transcription of the lac operon in *Escherichia coli* (*E. coli*) [Raibaud *et al.* (1984) *Annu. Rev. Genet.* 18:173]. Regulated expression may therefore be either positive or negative, thereby either enhancing or reducing transcription.

Sequences encoding metabolic pathway enzymes provide particularly useful promoter sequences. Examples include promoter sequences derived from sugar metabolizing enzymes, such as galactose, lactose (*lac*) [Chang *et al.* (1977) *Nature* 198:1056], and maltose. Additional examples include promoter sequences derived from biosynthetic enzymes such as tryptophan (*trp*) [Goeddel *et al.* (1980) *Nuc. Acids Res.* 8:4057; Yelverton *et al.* (1981) *Nucl. Acids Res.* 9:731; US patent 4,738,921; EP-A-0036776 and EP-A-0121775]. The *g*-laotamase (*bla*) promoter system [Weissmann (1981) "The cloning of interferon and other mistakes." In *Interferon 3* (ed. I. Gresser)], bacteriophage lambda PL [Shimatake *et al.* (1981) *Nature* 292:128] and T5 [US patent 4,689,406] promoter systems also provide useful promoter sequences.

In addition, synthetic promoters which do not occur in nature also function as bacterial promoters. For example, transcription activation sequences of one bacterial or bacteriophage promoter may be joined with the operon sequences of another bacterial or bacteriophage promoter, creating a synthetic hybrid promoter [US patent 4,551,433]. For example, the *tac* promoter is a hybrid *trp-lac* promoter comprised of both *trp* promoter and *lac* operon sequences that is regulated by the *lac* repressor [Amann *et al.* (1983) *Gene* 25:167; de Boer *et al.* (1983) *Proc. Natl. Acad. Sci.* 80:21]. Furthermore, a bacterial promoter can include naturally occurring promoters of non-bacterial origin that have the ability to bind bacterial RNA polymerase and initiate transcription. A naturally occurring promoter of non-bacterial origin can also be coupled with a compatible RNA polymerase to produce high levels of expression of some genes in prokaryotes. The bacteriophage T7 RNA polymerase/promoter system is an example of a coupled promoter system [Studier *et al.* (1986) *J. Mol. Biol.* 189:113; Tabor *et al.* (1985) *Proc Natl. Acad. Sci.* 82:1074]. In addition, a hybrid promoter can also be comprised of a bacteriophage promoter and an *E. coli* operator region (EPO-A-0 267 851).

In addition to a functioning promoter sequence, an efficient ribosome binding site is also useful for the expression of foreign genes in prokaryotes. In *E. coli*, the ribosome binding site is called the Shine-Dalgarno (SD) sequence and includes an initiation codon (ATG) and a sequence 3-9 nucleotides in length located 3-11 nucleotides upstream of the initiation codon [Shine *et al.* (1975) *Nature* 254:34]. The SD sequence is thought to promote binding of mRNA to the ribosome by the pairing of bases between the SD sequence and the 3' end of *E. coli* 16S rRNA [Steitz *et al.* (1979) "Genetic signals and nucleotide sequences in messenger RNA." In *Biological*

*Regulation and Development: Gene Expression* (ed. R.F. Goldberger)]. To express eukaryotic genes and prokaryotic genes with weak ribosome-binding site [Sambrook *et al.* (1989) "Expression of cloned genes in *Escherichia coli*." In *Molecular Cloning: A Laboratory Manual*].

5 A DNA molecule may be expressed intracellularly. A promoter sequence may be directly linked with the DNA molecule, in which case the first amino acid at the N-terminus will always be a methionine, which is encoded by the ATG start codon. If desired, methionine at the N-terminus may be cleaved from the protein by *in vitro* incubation with cyanogen bromide or by either *in vivo* or *in vitro* incubation with a bacterial methionine N-terminal peptidase (EPO-A-0 219 237).

10 Fusion proteins provide an alternative to direct expression. Usually, a DNA sequence encoding the N-terminal portion of an endogenous bacterial protein, or other stable protein, is fused to the 5' end of heterologous coding sequences. Upon expression, this construct will provide a fusion of the two amino acid sequences. For example, the bacteriophage lambda cell gene can be linked at the 5' terminus of a foreign gene and expressed in bacteria. The resulting fusion protein preferably retains a site for a processing enzyme (factor Xa) to cleave the bacteriophage protein from the foreign gene [Nagai *et al.* (1984) *Nature* 309:810]. Fusion proteins can also be  
15 made with sequences from the *lacZ* [Jia *et al.* (1987) *Gene* 60:197], *trpE* [Allen *et al.* (1987) *J. Biotechnol.* 5:93; Makoff *et al.* (1989) *J. Gen. Microbiol.* 135:11], and *Chey* [EP-A-0 324 647] genes. The DNA sequence at the junction of the two amino acid sequences may or may not encode a cleavable site. Another example is a ubiquitin fusion protein. Such a fusion protein is made with the ubiquitin region that preferably retains a site for a processing enzyme (*e.g.* ubiquitin specific processing-protease) to cleave the ubiquitin from the foreign  
20 protein. Through this method, native foreign protein can be isolated [Miller *et al.* (1989) *Bio/Technology* 7:698].

Alternatively, foreign proteins can also be secreted from the cell by creating chimeric DNA molecules that encode a fusion protein comprised of a signal peptide sequence fragment that provides for secretion of the foreign protein in bacteria [US patent 4,336,336]. The signal sequence fragment usually encodes a signal peptide comprised of hydrophobic amino acids which direct the secretion of the protein from the cell. The protein is  
25 either secreted into the growth media (gram-positive bacteria) or into the periplasmic space, located between the inner and outer membrane of the cell (gram-negative bacteria). Preferably there are processing sites, which can be cleaved either *in vivo* or *in vitro* encoded between the signal peptide fragment and the foreign gene.

DNA encoding suitable signal sequences can be derived from genes for secreted bacterial proteins, such as the *E. coli* outer membrane protein gene (*ompA*) [Masui *et al.* (1983), in: *Experimental Manipulation of Gene*  
30 *Expression*; Ghayeb *et al.* (1984) *EMBO J.* 3:2437] and the *E. coli* alkaline phosphatase signal sequence (*phoA*) [Oka *et al.* (1985) *Proc. Natl. Acad. Sci.* 82:7212]. As an additional example, the signal sequence of the alpha-amylase gene from various *Bacillus* strains can be used to secrete heterologous proteins from *B. subtilis* [Palva *et al.* (1982) *Proc. Natl. Acad. Sci. USA* 79:5582; EP-A-0 244 042].

Usually, transcription termination sequences recognized by bacteria are regulatory regions located 3' to the  
35 translation stop codon, and thus together with the promoter flank the coding sequence. These sequences direct the transcription of an mRNA which can be translated into the polypeptide encoded by the DNA. Transcription termination sequences frequently include DNA sequences of about 50 nucleotides capable of forming stem loop structures that aid in terminating transcription. Examples include transcription termination sequences derived from genes with strong promoters, such as the *trp* gene in *E. coli* as well as other biosynthetic genes.

Usually, the above described components, comprising a promoter, signal sequence (if desired), coding sequence of interest, and transcription termination sequence, are put together into expression constructs. Expression constructs are often maintained in a replicon, such as an extrachromosomal element (e.g. plasmids) capable of stable maintenance in a host, such as bacteria. The replicon will have a replication system, thus allowing it to be maintained in a prokaryotic host either for expression or for cloning and amplification. In addition, a replicon may be either a high or low copy number plasmid. A high copy number plasmid will generally have a copy number ranging from about 5 to about 200, and usually about 10 to about 150. A host containing a high copy number plasmid will preferably contain at least about 10, and more preferably at least about 20 plasmids. Either a high or low copy number vector may be selected, depending upon the effect of the vector and the foreign protein on the host.

Alternatively, the expression constructs can be integrated into the bacterial genome with an integrating vector. Integrating vectors usually contain at least one sequence homologous to the bacterial chromosome that allows the vector to integrate. Integrations appear to result from recombinations between homologous DNA in the vector and the bacterial chromosome. For example, integrating vectors constructed with DNA from various *Bacillus* strains integrate into the *Bacillus* chromosome (EP-A- 0 127 328). Integrating vectors may also be comprised of bacteriophage or transposon sequences.

Usually, extrachromosomal and integrating expression constructs may contain selectable markers to allow for the selection of bacterial strains that have been transformed. Selectable markers can be expressed in the bacterial host and may include genes which render bacteria resistant to drugs such as ampicillin, chloramphenicol, erythromycin, kanamycin (neomycin), and tetracycline [Davies *et al.* (1978) *Annu. Rev. Microbiol.* 32:469]. Selectable markers may also include biosynthetic genes, such as those in the histidine, tryptophan, and leucine biosynthetic pathways.

Alternatively, some of the above described components can be put together in transformation vectors. Transformation vectors are usually comprised of a selectable market that is either maintained in a replicon or developed into an integrating vector, as described above.

Expression and transformation vectors, either extra-chromosomal replicons or integrating vectors, have been developed for transformation into many bacteria. For example, expression vectors have been developed for, *inter alia*, the following bacteria: *Bacillus subtilis* [Palva *et al.* (1982) *Proc. Natl. Acad. Sci. USA* 79:5582; EP-A-0 036 259 and EP-A-0 063 953; WO 84/04541], *Escherichia coli* [Shimatake *et al.* (1981) *Nature* 292:128; Amann *et al.* (1985) *Gene* 40:183; Studier *et al.* (1986) *J. Mol. Biol.* 189:113; EP-A-0 036 776, EP-A-0 136 829 and EP-A-0 136 907], *Streptococcus cremoris* [Powell *et al.* (1988) *Appl. Environ. Microbiol.* 54:655]; *Streptococcus lividans* [Powell *et al.* (1988) *Appl. Environ. Microbiol.* 54:655], *Streptomyces lividans* [US patent 4,745,056].

Methods of introducing exogenous DNA into bacterial hosts are well-known in the art, and usually include either the transformation of bacteria treated with  $\text{CaCl}_2$  or other agents, such as divalent cations and DMSO. DNA can also be introduced into bacterial cells by electroporation. Transformation procedures usually vary with the bacterial species to be transformed. See e.g. [Masson *et al.* (1989) *FEMS Microbiol. Lett.* 60:273; Palva *et al.* (1982) *Proc. Natl. Acad. Sci. USA* 79:5582; EP-A-0 036 259 and EP-A-0 063 953; WO 84/04541, *Bacillus*], [Miller *et al.* (1988) *Proc. Natl. Acad. Sci.* 85:856; Wang *et al.* (1990) *J. Bacteriol.* 172:949, *Campylobacter*], [Cohen *et al.* (1973) *Proc. Natl. Acad. Sci.* 69:2110; Dower *et al.* (1988) *Nucleic Acids Res.* 16:6127; Kushner (1978) "An improved method for transformation of *Escherichia coli* with ColE1-derived plasmids. In *Genetic*

*Engineering: Proceedings of the International Symposium on Genetic Engineering* (eds. H.W. Boyer and S. Nicosia); Mandel *et al.* (1970) *J. Mol. Biol.* 53:159; Taketo (1988) *Biochim. Biophys. Acta* 949:318; Escherichia], [Chassy *et al.* (1987) *FEMS Microbiol. Lett.* 44:173 Lactobacillus]; [Fiedler *et al.* (1988) *Anal. Biochem* 170:38, Pseudomonas]; [Augustin *et al.* (1990) *FEMS Microbiol. Lett.* 66:203, Staphylococcus],  
 5 [Barany *et al.* (1980) *J. Bacteriol.* 144:698; Harlander (1987) "Transformation of Streptococcus lactis by electroporation, in: *Streptococcal Genetics* (ed. J. Ferretti and R. Curtiss III); Perry *et al.* (1981) *Infect. Immun.* 32:1295; Powell *et al.* (1988) *Appl. Environ. Microbiol.* 54:655; Somkuti *et al.* (1987) *Proc. 4th Eur. Cong. Biotechnology* 1:412, Streptococcus].

#### v. Yeast Expression

10 Yeast expression systems are also known to one of ordinary skill in the art. A yeast promoter is any DNA sequence capable of binding yeast RNA polymerase and initiating the downstream (3') transcription of a coding sequence (e.g. structural gene) into mRNA. A promoter will have a transcription initiation region which is usually placed proximal to the 5' end of the coding sequence. This transcription initiation region usually includes an RNA polymerase binding site (the "TATA Box") and a transcription initiation site. A yeast promoter may  
 15 also have a second domain called an upstream activator sequence (UAS), which, if present, is usually distal to the structural gene. The UAS permits regulated (inducible) expression. Constitutive expression occurs in the absence of a UAS. Regulated expression may be either positive or negative, thereby either enhancing or reducing transcription.

20 Yeast is a fermenting organism with an active metabolic pathway, therefore sequences encoding enzymes in the metabolic pathway provide particularly useful promoter sequences. Examples include alcohol dehydrogenase (ADH) (EP-A-0 284 044), enolase, glucokinase, glucose-6-phosphate isomerase, glyceraldehyde-3-phosphate-dehydrogenase (GAP or GAPDH), hexokinase, phosphofructokinase, 3-phosphoglycerate mutase, and pyruvate kinase (PyK) (EPO-A-0 329 203). The yeast *PHO5* gene, encoding acid phosphatase, also provides useful promoter sequences [Myanohara *et al.* (1983) *Proc. Natl. Acad. Sci. USA* 80:1].

25 In addition, synthetic promoters which do not occur in nature also function as yeast promoters. For example, UAS sequences of one yeast promoter may be joined with the transcription activation region of another yeast promoter, creating a synthetic hybrid promoter. Examples of such hybrid promoters include the ADH regulatory sequence linked to the GAP transcription activation region (US Patent Nos. 4,876,197 and 4,880,734). Other examples of hybrid promoters include promoters which consist of the regulatory sequences of either the *ADH2*,  
 30 *GAL4*, *GAL10*, OR *PHO5* genes, combined with the transcriptional activation region of a glycolytic enzyme gene such as GAP or PyK (EP-A-0 164 556). Furthermore, a yeast promoter can include naturally occurring promoters of non-yeast origin that have the ability to bind yeast RNA polymerase and initiate transcription. Examples of such promoters include, *inter alia*, [Cohen *et al.* (1980) *Proc. Natl. Acad. Sci. USA* 77:1078; Henikoff *et al.* (1981) *Nature* 283:835; Hollenberg *et al.* (1981) *Curr. Topics Microbiol. Immunol.* 96:119;  
 35 Hollenberg *et al.* (1979) "The Expression of Bacterial Antibiotic Resistance Genes in the Yeast *Saccharomyces cerevisiae*," in: *Plasmids of Medical, Environmental and Commercial Importance* (eds. K.N. Timmis and A. Puhler); Mercerau-Puigalon *et al.* (1980) *Gene* 11:163; Panthier *et al.* (1980) *Curr. Genet.* 2:109;].

A DNA molecule may be expressed intracellularly in yeast. A promoter sequence may be directly linked with the DNA molecule, in which case the first amino acid at the N-terminus of the recombinant protein will always

be a methionine, which is encoded by the ATG start codon. If desired, methionine at the N-terminus may be cleaved from the protein by *in vitro* incubation with cyanogen bromide.

Fusion proteins provide an alternative for yeast expression systems, as well as in mammalian, baculovirus, and bacterial expression systems. Usually, a DNA sequence encoding the N-terminal portion of an endogenous yeast protein, or other stable protein, is fused to the 5' end of heterologous coding sequences. Upon expression, this construct will provide a fusion of the two amino acid sequences. For example, the yeast or human superoxide dismutase (SOD) gene, can be linked at the 5' terminus of a foreign gene and expressed in yeast. The DNA sequence at the junction of the two amino acid sequences may or may not encode a cleavable site. See *e.g.* EP-A-0 196 056. Another example is a ubiquitin fusion protein. Such a fusion protein is made with the ubiquitin region that preferably retains a site for a processing enzyme (*e.g.* ubiquitin-specific processing protease) to cleave the ubiquitin from the foreign protein. Through this method, therefore, native foreign protein can be isolated (*e.g.* WO88/024066).

Alternatively, foreign proteins can also be secreted from the cell into the growth media by creating chimeric DNA molecules that encode a fusion protein comprised of a leader sequence fragment that provide for secretion in yeast of the foreign protein. Preferably, there are processing sites encoded between the leader fragment and the foreign gene that can be cleaved either *in vivo* or *in vitro*. The leader sequence fragment usually encodes a signal peptide comprised of hydrophobic amino acids which direct the secretion of the protein from the cell.

DNA encoding suitable signal sequences can be derived from genes for secreted yeast proteins, such as the genes for invertase (EP-A-0012873; JPO 62,096,086) and A-factor (US patent 4,588,684). Alternatively, leaders of non-yeast origin exist, such as an interferon leader, that also provide for secretion in yeast (EP-A-0060057).

A preferred class of secretion leaders are those that employ a fragment of the yeast alpha-factor gene, which contains both a "pre" signal sequence, and a "pro" region. The types of alpha-factor fragments that can be employed include the full-length pre-pro alpha factor leader (about 83 amino acid residues) as well as truncated alpha-factor leaders (usually about 25 to about 50 amino acid residues) (US Patents 4,546,083 and 4,870,008; EP-A-0 324 274). Additional leaders employing an alpha-factor leader fragment that provides for secretion include hybrid alpha-factor leaders made with a presequence of a first yeast, but a pro-region from a second yeast alphafactor. (*e.g.* see WO 89/02463.)

Usually, transcription termination sequences recognized by yeast are regulatory regions located 3' to the translation stop codon, and thus together with the promoter flank the coding sequence. These sequences direct the transcription of an mRNA which can be translated into the polypeptide encoded by the DNA. Examples of transcription terminator sequence and other yeast-recognized termination sequences, such as those coding for glycolytic enzymes.

Usually, the above described components, comprising a promoter, leader (if desired), coding sequence of interest, and transcription termination sequence, are put together into expression constructs. Expression constructs are often maintained in a replicon, such as an extrachromosomal element (*e.g.* plasmids) capable of stable maintenance in a host, such as yeast or bacteria. The replicon may have two replication systems, thus allowing it to be maintained, for example, in yeast for expression and in a prokaryotic host for cloning and amplification. Examples of such yeast-bacteria shuttle vectors include YE24 [Botstein *et al.* (1979) *Gene* 8:17-24], pCI/1 [Brake *et al.* (1984) *Proc. Natl. Acad. Sci USA* 81:4642-4646], and YRp17 [Stinchcomb *et al.* (1982) *J. Mol. Biol.* 158:157]. In addition, a replicon may be either a high or low copy number plasmid. A high copy



number plasmid will generally have a copy number ranging from about 5 to about 200, and usually about 10 to about 150. A host containing a high copy number plasmid will preferably have at least about 10, and more preferably at least about 20. Enter a high or low copy number vector may be selected, depending upon the effect of the vector and the foreign protein on the host. See *e.g.* Brake *et al.*, *supra*.

- 5 Alternatively, the expression constructs can be integrated into the yeast genome with an integrating vector. Integrating vectors usually contain at least one sequence homologous to a yeast chromosome that allows the vector to integrate, and preferably contain two homologous sequences flanking the expression construct. Integrations appear to result from recombinations between homologous DNA in the vector and the yeast chromosome [Orr-Weaver *et al.* (1983) *Methods in Enzymol.* 101:228-245]. An integrating vector may be  
10 directed to a specific locus in yeast by selecting the appropriate homologous sequence for inclusion in the vector. See Orr-Weaver *et al.*, *supra*. One or more expression construct may integrate, possibly affecting levels of recombinant protein produced [Rine *et al.* (1983) *Proc. Natl. Acad. Sci. USA* 80:6750]. The chromosomal sequences included in the vector can occur either as a single segment in the vector, which results in the integration of the entire vector, or two segments homologous to adjacent segments in the chromosome and flanking the  
15 expression construct in the vector, which can result in the stable integration of only the expression construct.

Usually, extrachromosomal and integrating expression constructs may contain selectable markers to allow for the selection of yeast strains that have been transformed. Selectable markers may include biosynthetic genes that can be expressed in the yeast host, such as *ADE2*, *HIS4*, *LEU2*, *TRP1*, and *ALG7*, and the G418 resistance gene, which confer resistance in yeast cells to tunicamycin and G418, respectively. In addition, a suitable selectable  
20 marker may also provide yeast with the ability to grow in the presence of toxic compounds, such as metal. For example, the presence of *CUP1* allows yeast to grow in the presence of copper ions [Butt *et al.* (1987) *Microbiol. Rev.* 51:351].

Alternatively, some of the above described components can be put together into transformation vectors. Transformation vectors are usually comprised of a selectable marker that is either maintained in a replicon or  
25 developed into an integrating vector, as described above.

Expression and transformation vectors, either extrachromosomal replicons or integrating vectors, have been developed for transformation into many yeasts. For example, expression vectors have been developed for, *inter alia*, the following yeasts: *Candida albicans* [Kurtz, *et al.* (1986) *Mol. Cell. Biol.* 6:142], *Candida maltosa* [Kunze, *et al.* (1985) *J. Basic Microbiol.* 25:141], *Hansenula polymorpha* [Gleeson, *et al.* (1986) *J. Gen. Microbiol.* 132:3459; Roggenkamp *et al.* (1986) *Mol. Gen. Genet.* 202:302], *Kluyveromyces fragilis* [Das, *et al.*  
30 (1984) *J. Bacteriol.* 158:1165], *Kluyveromyces lactis* [De Louvencourt *et al.* (1983) *J. Bacteriol.* 154:737; Van den Berg *et al.* (1990) *Bio/Technology* 8:135], *Pichia guilliermondii* [Kunze *et al.* (1985) *J. Basic Microbiol.* 25:141], *Pichia pastoris* [Cregg, *et al.* (1985) *Mol. Cell. Biol.* 5:3376; US Patent Nos. 4,837,148 and 4,929,555], *Saccharomyces cerevisiae* [Hinnen *et al.* (1978) *Proc. Natl. Acad. Sci. USA* 75:1929; Ito *et al.* (1983) *J. Bacteriol.* 153:163], *Schizosaccharomyces pombe* [Beach and Nurse (1981) *Nature* 300:706], and *Yarrowia lipolytica* [Davidow, *et al.* (1985) *Curr. Genet.* 10:380471 Gaillardin, *et al.* (1985) *Curr. Genet.* 10:49].

Methods of introducing exogenous DNA into yeast hosts are well-known in the art, and usually include either the transformation of spheroplasts or of intact yeast cells treated with alkali cations. Transformation procedures usually vary with the yeast species to be transformed. See *e.g.* [Kurtz *et al.* (1986) *Mol. Cell. Biol.* 6:142; Kunze  
40 *et al.* (1985) *J. Basic Microbiol.* 25:141; *Candida*]; [Gleeson *et al.* (1986) *J. Gen. Microbiol.* 132:3459;

Roggkamp *et al.* (1986) *Mol. Gen. Genet.* 202:302; Hansen *et al.* (1984) *J. Bacteriol.* 158:1165; De Louvencourt *et al.* (1983) *J. Bacteriol.* 154:1165; Van den Berg *et al.* (1990) *Bio/Technology* 8:135; Kluyveromyces]; [Cregg *et al.* (1985) *Mol. Cell. Biol.* 5:3376; Kunze *et al.* (1985) *J. Basic Microbiol.* 25:141; US Patents 4,837,148 & 4,929,555; Pichia]; [Hinnen *et al.* (1978) *Proc. Natl. Acad. Sci. USA* 75:1929; Ito *et al.* (1983) *J. Bacteriol.* 153:163 *Saccharomyces*]; [Beach & Nurse (1981) *Nature* 300:706; *Schizosaccharomyces*]; [Davidow *et al.* (1985) *Curr. Genet.* 10:39; Gaillardin *et al.* (1985) *Curr. Genet.* 10:49; *Yarrowia*].

#### Pharmaceutical Compositions

Pharmaceutical compositions can comprise polypeptides and/or nucleic acid of the invention. The pharmaceutical compositions will comprise a therapeutically effective amount of either polypeptides, antibodies, or polynucleotides of the claimed invention.

The term "therapeutically effective amount" as used herein refers to an amount of a therapeutic agent to treat, ameliorate, or prevent a desired disease or condition, or to exhibit a detectable therapeutic or preventative effect. The effect can be detected by, for example, chemical markers or antigen levels. Therapeutic effects also include reduction in physical symptoms, such as decreased body temperature. The precise effective amount for a subject will depend upon the subject's size and health, the nature and extent of the condition, and the therapeutics or combination of therapeutics selected for administration. Thus, it is not useful to specify an exact effective amount in advance. However, the effective amount for a given situation can be determined by routine experimentation and is within the judgement of the clinician.

For purposes of the present invention, an effective dose will be from about 0.01 mg/ kg to 50 mg/kg or 0.05 mg/kg to about 10 mg/kg of the DNA constructs in the individual to which it is administered.

A pharmaceutical composition can also contain a pharmaceutically acceptable carrier. The term "pharmaceutically acceptable carrier" refers to a carrier for administration of a therapeutic agent, such as antibodies or a polypeptide, genes, and other therapeutic agents. The term refers to any pharmaceutical carrier that does not itself induce the production of antibodies harmful to the individual receiving the composition, and which may be administered without undue toxicity. Suitable carriers may be large, slowly metabolized macromolecules such as proteins, polysaccharides, polylactic acids, polyglycolic acids, polymeric amino acids, amino acid copolymers, and inactive virus particles. Such carriers are well known to those of ordinary skill in the art.

Pharmaceutically acceptable salts can be used therein, for example, mineral acid salts such as hydrochlorides, hydrobromides, phosphates, sulfates, and the like; and the salts of organic acids such as acetates, propionates, malonates, benzoates, and the like. A thorough discussion of pharmaceutically acceptable excipients is available in Remington's Pharmaceutical Sciences (Mack Pub. Co., N.J. 1991).

Pharmaceutically acceptable carriers in therapeutic compositions may contain liquids such as water, saline, glycerol and ethanol. Additionally, auxiliary substances, such as wetting or emulsifying agents, pH buffering substances, and the like, may be present in such vehicles. Typically, the therapeutic compositions are prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid vehicles prior to injection may also be prepared. Liposomes are included within the definition of a pharmaceutically acceptable carrier.

Delivery Methods

Once formulated, the compositions of the invention can be administered directly to the subject. The subjects to be treated can be animals; in particular, human subjects can be treated.

5 Direct delivery of the compositions will generally be accomplished by injection, either subcutaneously, intraperitoneally, intravenously or intramuscularly or delivered to the interstitial space of a tissue. The compositions can also be administered into a lesion. Other modes of administration include oral and pulmonary administration, suppositories, and transdermal or transcutaneous applications (e.g. see WO98/20734), needles, and gene guns or hyposprays. Dosage treatment may be a single dose schedule or a multiple dose schedule.

Vaccines

10 Vaccines according to the invention may either be prophylactic (ie. to prevent infection) or therapeutic (ie. to treat disease after infection).

Such vaccines comprise immunising antigen(s), immunogen(s), polypeptide(s), protein(s) or nucleic acid, usually in combination with "pharmaceutically acceptable carriers," which include any carrier that does not itself induce the production of antibodies harmful to the individual receiving the composition. Suitable carriers are typically large, slowly metabolized macromolecules such as proteins, polysaccharides, polylactic acids, 15 polyglycolic acids, polymeric amino acids, amino acid copolymers, lipid aggregates (such as oil droplets or liposomes), and inactive virus particles. Such carriers are well known to those of ordinary skill in the art. Additionally, these carriers may function as immunostimulating agents ("adjuvants"). Furthermore, the antigen or immunogen may be conjugated to a bacterial toxoid, such as a toxoid from diphtheria, tetanus, cholera, *H. pylori*, etc. pathogens.

Preferred adjuvants to enhance effectiveness of the composition include, but are not limited to: (1) aluminum salts (alum), such as aluminum hydroxide, aluminum phosphate, aluminum sulfate, etc; (2) oil-in-water emulsion formulations (with or without other specific immunostimulating agents such as muramyl peptides (see below) or bacterial cell wall components), such as for example (a) MF59™ (WO 90/14837; Chapter 10 in 25 *Vaccine design: the subunit and adjuvant approach*, eds. Powell & Newman, Plenum Press 1995), containing 5% Squalene, 0.5% Tween 80, and 0.5% Span 85 (optionally containing various amounts of MTP-PE (see below), although not required) formulated into submicron particles using a microfluidizer such as Model 110Y microfluidizer (Microfluidics, Newton, MA), (b) SAF, containing 10% Squalene, 0.4% Tween 80, 5% pluronic-blocked polymer L121, and thr-MDP (see below) either microfluidized into a submicron emulsion or vortexed to generate a larger particle size emulsion, and (c) Ribi™ adjuvant system (RAS), (Ribi Immunochem, Hamilton, 30 MT) containing 2% Squalene, 0.2% Tween 80, and one or more bacterial cell wall components from the group consisting of monophosphorylipid A (MPL), trehalose dimycolate (TDM), and cell wall skeleton (CWS), preferably MPL + CWS (Detox™); (3) saponin adjuvants, such as Stimulon™ (Cambridge Bioscience, Worcester, MA) may be used or particles generated therefrom such as ISCOMs (immunostimulating complexes); (4) Complete Freund's Adjuvant (CFA) and Incomplete Freund's Adjuvant (IFA); (5) cytokines, 35 such as interleukins (e.g. IL-1, IL-2, IL-4, IL-5, IL-6, IL-7, IL-12, etc.), interferons (e.g. gamma interferon), macrophage colony stimulating factor (M-CSF), tumor necrosis factor (TNF), etc; and (6) other substances that act as immunostimulating agents to enhance the effectiveness of the composition. Alum and MF59™ are preferred.

As mentioned above, muramyl peptides include, but are not limited to, N-acetyl-muramyl-L-threonyl-D-isoglutamine (thr-MDP), N-acetyl-normuramyl-L-alanyl-D-isoglutamine (nor-MDP), N-acetylmuramyl-L-alanyl-D-isoglutaminyl-L-alanine-2-(1'-2'-dipalmitoyl-sn-glycero-3-hydroxyphosphoryloxy)-ethylamine (MTP-PE), etc.

5 The immunogenic compositions (e.g. the immunising antigen/immunogen/polypeptide/protein/ nucleic acid, pharmaceutically acceptable carrier, and adjuvant) typically will contain diluents, such as water, saline, glycerol, ethanol, etc. Additionally, auxiliary substances, such as wetting or emulsifying agents, pH buffering substances, and the like, may be present in such vehicles.

10 Typically, the immunogenic compositions are prepared as injectables, either as liquid solutions or suspensions; solid forms suitable for solution in, or suspension in, liquid vehicles prior to injection may also be prepared. The preparation also may be emulsified or encapsulated in liposomes for enhanced adjuvant effect, as discussed above under pharmaceutically acceptable carriers.

15 Immunogenic compositions used as vaccines comprise an immunologically effective amount of the antigenic or immunogenic polypeptides, as well as any other of the above-mentioned components, as needed. By "immunologically effective amount", it is meant that the administration of that amount to an individual, either in a single dose or as part of a series, is effective for treatment or prevention. This amount varies depending upon the health and physical condition of the individual to be treated, the taxonomic group of individual to be treated (e.g. nonhuman primate, primate, etc.), the capacity of the individual's immune system to synthesize antibodies, the degree of protection desired, the formulation of the vaccine, the treating doctor's assessment of the medical situation, and other relevant factors. It is expected that the amount will fall in a relatively broad range that can be  
20 determined through routine trials.

The immunogenic compositions are conventionally administered parenterally, e.g. by injection, either subcutaneously, intramuscularly, or transdermally/transcutaneously (e.g. WO98/20734). Additional formulations suitable for other modes of administration include oral and pulmonary formulations, suppositories, and transdermal applications. Dosage treatment may be a single dose schedule or a multiple dose schedule. The vaccine may be  
25 administered in conjunction with other immunoregulatory agents.

As an alternative to protein-based vaccines, DNA vaccination may be employed [e.g. Robinson & Torres (1997) *Seminars in Immunology* 9:271-283; Donnelly *et al.* (1997) *Annu Rev Immunol* 15:617-648; see later herein].

#### Gene Delivery Vehicles

30 Gene therapy vehicles for delivery of constructs including a coding sequence of a therapeutic of the invention, to be delivered to the mammal for expression in the mammal, can be administered either locally or systemically. These constructs can utilize viral or non-viral vector approaches in *in vivo* or *ex vivo* modality. Expression of such coding sequence can be induced using endogenous mammalian or heterologous promoters. Expression of the coding sequence *in vivo* can be either constitutive or regulated.

35 The invention includes gene delivery vehicles capable of expressing the contemplated nucleic acid sequences. The gene delivery vehicle is preferably a viral vector and, more preferably, a retroviral, adenoviral, adeno-associated viral (AAV), herpes viral, or alphavirus vector. The viral vector can also be an astrovirus, coronavirus, orthomyxovirus, papovavirus, paramyxovirus, parvovirus, picornavirus, poxvirus, or togavirus viral vector. See generally, Jolly (1994) *Cancer Gene Therapy* 1:51-64; Kimura (1994) *Human Gene Therapy* 5:845-852; Connelly (1995) *Human Gene Therapy* 6:185-193; and Kaplitt (1994) *Nature Genetics* 6:148-153.

Retroviral vectors are well known in the art and we contemplate that any retroviral gene therapy vector is employable in the invention, including B, C and D type retroviruses, xenotropic retroviruses (for example, NZB-X1, NZB-X2 and NZB9-1 (see O'Neill (1985) *J. Virol.* 53:160) polytropic retroviruses *e.g.* MCF and MCF-MLV (see Kelly (1983) *J. Virol.* 45:291), spumaviruses and lentiviruses. See RNA Tumor Viruses, 5 Second Edition, Cold Spring Harbor Laboratory, 1985.

Portions of the retroviral gene therapy vector may be derived from different retroviruses. For example, retrovector LTRs may be derived from a Murine Sarcoma Virus, a tRNA binding site from a Rous Sarcoma Virus, a packaging signal from a Murine Leukemia Virus, and an origin of second strand synthesis from an Avian Leukosis Virus.

10 These recombinant retroviral vectors may be used to generate transduction competent retroviral vector particles by introducing them into appropriate packaging cell lines (see US patent 5,591,624). Retrovirus vectors can be constructed for site-specific integration into host cell DNA by incorporation of a chimeric integrase enzyme into the retroviral particle (see WO96/37626). It is preferable that the recombinant viral vector is a replication defective recombinant virus.

15 Packaging cell lines suitable for use with the above-described retrovirus vectors are well known in the art, are readily prepared (see WO95/30763 and WO92/05266), and can be used to create producer cell lines (also termed vector cell lines or "VCLs") for the production of recombinant vector particles. Preferably, the packaging cell lines are made from human parent cells (*e.g.* HT1080 cells) or mink parent cell lines, which eliminates inactivation in human serum.

20 Preferred retroviruses for the construction of retroviral gene therapy vectors include Avian Leukosis Virus, Bovine Leukemia Virus, Murine Leukemia Virus, Mink-Cell Focus-Inducing Virus, Murine Sarcoma Virus, Reticuloendotheliosis Virus and Rous Sarcoma Virus. Particularly preferred Murine Leukemia Viruses include 4070A and 1504A (Hartley and Rowe (1976) *J Virol* 19:19-25), Abelson (ATCC No. VR-999), Friend (ATCC No. VR-245), Graffi, Gross (ATCC No. VR-590), Kirsten, Harvey Sarcoma Virus and Rauscher (ATCC No. 25 VR-998) and Moloney Murine Leukemia Virus (ATCC No. VR-190). Such retroviruses may be obtained from depositories or collections such as the American Type Culture Collection ("ATCC") in Rockville, Maryland or isolated from known sources using commonly available techniques.

Exemplary known retroviral gene therapy vectors employable in this invention include those described in patent applications GB2200651, EP0415731, EP0345242, EP0334301, WO89/02468; WO89/05349, WO89/09271, 30 WO90/02806, WO90/07936, WO94/03622, WO93/25698, WO93/25234, WO93/11230, WO93/10218, WO91/02805, WO91/02825, WO95/07994, US 5,219,740, US 4,405,712, US 4,861,719, US 4,980,289, US 4,777,127, US 5,591,624. See also Vile (1993) *Cancer Res* 53:3860-3864; Vile (1993) *Cancer Res* 53:962-967; Ram (1993) *Cancer Res* 53 (1993) 83-88; Takamiya (1992) *J Neurosci Res* 33:493-503; Baba (1993) *J Neurosurg* 79:729-735; Mann (1983) *Cell* 33:153; Cane (1984) *Proc Natl Acad Sci* 81:6349; and Miller (1990) 35 *Human Gene Therapy* 1.

Human adenoviral gene therapy vectors are also known in the art and employable in this invention. See, for example, Berkner (1988) *Biotechniques* 6:616 and Rosenfeld (1991) *Science* 252:431, and WO93/07283, WO93/06223, and WO93/07282. Exemplary known adenoviral gene therapy vectors employable in this invention include those described in the above referenced documents and in WO94/12649, WO93/03769, 40 WO93/19191, WO94/28938, WO95/11984, WO95/00655, WO95/27071, WO95/29993, WO95/34671,

WO96/05320, WO94/08026, WO94/11506, WO93/06223, WO94/24299, WO95/14102, WO95/24297, WO95/02697, WO94/28152, WO94/24299, WO95/09241, WO95/25807, WO95/05835, WO94/18922 and WO95/09654. Alternatively, administration of DNA linked to killed adenovirus as described in Curiel (1992) *Hum. Gene Ther.* 3:147-154 may be employed. The gene delivery vehicles of the invention also include

5 adenovirus associated virus (AAV) vectors. Leading and preferred examples of such vectors for use in this invention are the AAV-2 based vectors disclosed in Srivastava, WO93/09239. Most preferred AAV vectors comprise the two AAV inverted terminal repeats in which the native D-sequences are modified by substitution of nucleotides, such that at least 5 native nucleotides and up to 18 native nucleotides, preferably at least 10 native nucleotides up to 18 native nucleotides, most preferably 10 native nucleotides are retained and the

10 remaining nucleotides of the D-sequence are deleted or replaced with non-native nucleotides. The native D-sequences of the AAV inverted terminal repeats are sequences of 20 consecutive nucleotides in each AAV inverted terminal repeat (*ie.* there is one sequence at each end) which are not involved in HP formation. The non-native replacement nucleotide may be any nucleotide other than the nucleotide found in the native D-sequence in the same position. Other employable exemplary AAV vectors are pWP-19, pWN-1, both of

15 which are disclosed in Nahreini (1993) *Gene* 124:257-262. Another example of such an AAV vector is psub201 (see Samulski (1987) *J. Virol.* 61:3096). Another exemplary AAV vector is the Double-D ITR vector. Construction of the Double-D ITR vector is disclosed in US Patent 5,478,745. Still other vectors are those disclosed in Carter US Patent 4,797,368 and Muzyczka US Patent 5,139,941, Chartejee US Patent 5,474,935, and Kotin WO94/288157. Yet a further example of an AAV vector employable in this invention is

20 SSV9AFABTKneo, which contains the AFP enhancer and albumin promoter and directs expression predominantly in the liver. Its structure and construction are disclosed in Su (1996) *Human Gene Therapy* 7:463-470. Additional AAV gene therapy vectors are described in US 5,354,678, US 5,173,414, US 5,139,941, and US 5,252,479.

The gene therapy vectors of the invention also include herpes vectors. Leading and preferred examples are

25 herpes simplex virus vectors containing a sequence encoding a thymidine kinase polypeptide such as those disclosed in US 5,288,641 and EP0176170 (Roizman). Additional exemplary herpes simplex virus vectors include HFEM/ICP6-LacZ disclosed in WO95/04139 (Wistar), pHSVlac described in Geller (1988) *Science* 241:1667-1669 and in WO90/09441 & WO92/07945, HSV Us3::pgC-lacZ described in Fink (1992) *Human Gene Therapy* 3:11-19 and HSV 7134, 2 RH 105 and GAL4 described in EP 0453242 (Breakefield), and those

30 deposited with ATCC as accession numbers ATCC VR-977 and ATCC VR-260.

Also contemplated are alpha virus gene therapy vectors that can be employed in this invention. Preferred alpha virus vectors are Sindbis viruses vectors. Togaviruses, Semliki Forest virus (ATCC VR-67; ATCC VR-1247), Middleberg virus (ATCC VR-370), Ross River virus (ATCC VR-373; ATCC VR-1246), Venezuelan equine encephalitis virus (ATCC VR923; ATCC VR-1250; ATCC VR-1249; ATCC VR-532), and those described in

35 US patents 5,091,309, 5,217,879, and WO92/10578. More particularly, those alpha virus vectors described in US Serial No. 08/405,627, filed March 15, 1995, WO94/21792, WO92/10578, WO95/07994, US 5,091,309 and US 5,217,879 are employable. Such alpha viruses may be obtained from depositories or collections such as the ATCC in Rockville, Maryland or isolated from known sources using commonly available techniques. Preferably, alphavirus vectors with reduced cytotoxicity are used (see USSN 08/679640).

40 DNA vector systems such as eukaryotic layered expression systems are also useful for expressing the nucleic acids of the invention. See WO95/07994 for a detailed description of eukaryotic layered expression systems.

Preferably, the eukaryotic layered expression systems of the invention are derived from alphavirus vectors and most preferably from Sindbis viral vectors.

Other viral vectors suitable for use in the present invention include those derived from poliovirus, for example ATCC VR-58 and those described in Evans, *Nature* 339 (1989) 385 and Sabin (1973) *J. Biol. Standardization* 1:115; rhinovirus, for example ATCC VR-1110 and those described in Arnold (1990) *J Cell Biochem* L401; pox viruses such as canary pox virus or vaccinia virus, for example ATCC VR-111 and ATCC VR-2010 and those described in Fisher-Hoch (1989) *Proc Natl Acad Sci* 86:317; Flexner (1989) *Ann NY Acad Sci* 569:86, Flexner (1990) *Vaccine* 8:17; in US 4,603,112 and US 4,769,330 and WO89/01973; SV40 virus, for example ATCC VR-305 and those described in Mulligan (1979) *Nature* 277:108 and Madzak (1992) *J Gen Virol* 73:1533; influenza virus, for example ATCC VR-797 and recombinant influenza viruses made employing reverse genetics techniques as described in US 5,166,057 and in Enami (1990) *Proc Natl Acad Sci* 87:3802-3805; Enami & Palese (1991) *J Virol* 65:2711-2713 and Luytjes (1989) *Cell* 59:110, (see also McMichael (1983) *NEJ Med* 309:13, and Yap (1978) *Nature* 273:238 and *Nature* (1979) 277:108); human immunodeficiency virus as described in EP-0386882 and in Buchschacher (1992) *J. Virol.* 66:2731; measles virus, for example ATCC VR-67 and VR-1247 and those described in EP-0440219; Aura virus, for example ATCC VR-368; Bebaru virus, for example ATCC VR-600 and ATCC VR-1240; Cabassou virus, for example ATCC VR-922; Chikungunya virus, for example ATCC VR-64 and ATCC VR-1241; Fort Morgan Virus, for example ATCC VR-924; Getah virus, for example ATCC VR-369 and ATCC VR-1243; Kyzylagach virus, for example ATCC VR-927; Mayaro virus, for example ATCC VR-66; Mucambo virus, for example ATCC VR-580 and ATCC VR-1244; Ndumu virus, for example ATCC VR-371; Pixuna virus, for example ATCC VR-372 and ATCC VR-1245; Tonate virus, for example ATCC VR-925; Trinit virus, for example ATCC VR-469; Una virus, for example ATCC VR-374; Whataroa virus, for example ATCC VR-926; Y-62-33 virus, for example ATCC VR-375; O'Nyong virus, Eastern encephalitis virus, for example ATCC VR-65 and ATCC VR-1242; Western encephalitis virus, for example ATCC VR-70; ATCC VR-1251, ATCC VR-622 and ATCC VR-1252; and coronavirus, for example ATCC VR-740 and those described in Hamre (1966) *Proc Soc Exp Biol Med* 121:190.

Delivery of the compositions of this invention into cells is not limited to the above mentioned viral vectors. Other delivery methods and media may be employed such as, for example, nucleic acid expression vectors, polycationic condensed DNA linked or unlinked to killed adenovirus alone, for example see US Serial No. 08/366,787, filed December 30, 1994 and Curiel (1992) *Hum Gene Ther* 3:147-154 ligand linked DNA, for example see Wu (1989) *J Biol Chem* 264:16985-16987, eucaryotic cell delivery vehicles cells, for example see US Serial No.08/240,030, filed May 9, 1994, and US Serial No. 08/404,796, deposition of photopolymerized hydrogel materials, hand-held gene transfer particle gun, as described in US Patent 5,149,655, ionizing radiation as described in US5,206,152 and in WO92/11033, nucleic charge neutralization or fusion with cell membranes. Additional approaches are described in Philip (1994) *Mol Cell Biol* 14:2411-2418 and in Woffendin (1994) *Proc Natl Acad Sci* 91:1581-1585.

Particle mediated gene transfer may be employed, for example see US Serial No. 60/023,867. Briefly, the sequence can be inserted into conventional vectors that contain conventional control sequences for high level expression, and then incubated with synthetic gene transfer molecules such as polymeric DNA-binding cations like polylysine, protamine, and albumin, linked to cell targeting ligands such as asialoorosomucoid, as described in Wu & Wu (1987) *J. Biol. Chem.* 262:4429-4432, insulin as described in Hucked (1990) *Biochem Pharmacol* 40:253-263, galactose as described in Plank (1992) *Bioconjugate Chem* 3:533-539, lactose or transferrin.

Naked DNA may also be employed. Exemplary naked DNA introduction methods are described in WO90/11092 and US 5,580,859. Uptake efficiency may be improved using biodegradable latex beads. DNA coated latex beads are efficiently transported into cells after endocytosis initiation by the beads. The method may be improved further by treatment of the beads to increase hydrophobicity and thereby facilitate disruption of the endosome and release of the DNA into the cytoplasm.

Liposomes that can act as gene delivery vehicles are described in US 5,422,120, WO95/13796, WO94/23697, WO91/14445 and EP-524,968. As described in USSN. 60/023,867, on non-viral delivery, the nucleic acid sequences encoding a polypeptide can be inserted into conventional vectors that contain conventional control sequences for high level expression, and then be incubated with synthetic gene transfer molecules such as polymeric DNA-binding cations like polylysine, protamine, and albumin, linked to cell targeting ligands such as asialoorosomucoid, insulin, galactose, lactose, or transferrin. Other delivery systems include the use of liposomes to encapsulate DNA comprising the gene under the control of a variety of tissue-specific or ubiquitously-active promoters. Further non-viral delivery suitable for use includes mechanical delivery systems such as the approach described in Woffendin *et al* (1994) *Proc. Natl. Acad. Sci. USA* 91(24):11581-11585.

Moreover, the coding sequence and the product of expression of such can be delivered through deposition of photopolymerized hydrogel materials. Other conventional methods for gene delivery that can be used for delivery of the coding sequence include, for example, use of hand-held gene transfer particle gun, as described in US 5,149,655; use of ionizing radiation for activating transferred gene, as described in US 5,206,152 and WO92/11033

Exemplary liposome and polycationic gene delivery vehicles are those described in US 5,422,120 and 4,762,915; in WO 95/13796; WO94/23697; and WO91/14445; in EP-0524968; and in Stryer, *Biochemistry*, pages 236-240 (1975) W.H. Freeman, San Francisco; Szoka (1980) *Biochem Biophys Acta* 600:1; Bayer (1979) *Biochem Biophys Acta* 550:464; Rivnay (1987) *Meth Enzymol* 149:119; Wang (1987) *Proc Natl Acad Sci* 84:7851; Plant (1989) *Anal Biochem* 176:420.

A polynucleotide composition can comprises therapeutically effective amount of a gene therapy vehicle, as the term is defined above. For purposes of the present invention, an effective dose will be from about 0.01 mg/ kg to 50 mg/kg or 0.05 mg/kg to about 10 mg/kg of the DNA constructs in the individual to which it is administered.

#### Delivery Methods

Once formulated, the polynucleotide compositions of the invention can be administered (1) directly to the subject; (2) delivered *ex vivo*, to cells derived from the subject; or (3) *in vitro* for recombinant protein expression. The subjects to be treated can be mammals or birds. Also, human subjects can be treated.

Direct delivery of the compositions will generally be accomplished by injection, either subcutaneously, intraperitoneally, intravenously or intramuscularly or delivered to the interstitial space of a tissue. The compositions can also be administered into a lesion. Other modes of administration include oral and pulmonary administration, suppositories, and transdermal or transcutaneous applications (*e.g.* see WO98/20734), needles, and gene guns or hyposprays. Dosage treatment may be a single dose schedule or a multiple dose schedule.

Methods for the *ex vivo* delivery and reimplantation of transformed cells into a subject are known in the art and described in *e.g.* WO93/14778. Examples of cells useful in *ex vivo* applications include, for example, stem cells, particularly hematopoietic, lymph cells, macrophages, dendritic cells, or tumor cells.



Generally, delivery of nucleic acids for both *ex vivo* and *in vitro* applications can be accomplished by the following procedures, for example, dextran-mediated transfection, calcium phosphate precipitation, polybrene mediated transfection, protoplast fusion, electroporation, encapsulation of the polynucleotide(s) in liposomes, and direct microinjection of the DNA into nuclei, all well known in the art.

5 Polynucleotide and polypeptide pharmaceutical compositions

In addition to the pharmaceutically acceptable carriers and salts described above, the following additional agents can be used with polynucleotide and/or polypeptide compositions.

A. Polypeptides

10 One example are polypeptides which include, without limitation: asialoglycoproteins; antibodies; antibody fragments; ferritin; interleukins; interferons, granulocyte, macrophage colony stimulating factor (GM-CSF), granulocyte colony stimulating factor (G-CSF), macrophage colony stimulating factor (M-CSF), stem cell factor and erythropoietin. Viral antigens, such as envelope proteins, can also be used. Also, proteins from other invasive organisms, such as the 17 amino acid peptide from the circumsporozoite protein of plasmodium falciparum known as RII.

15 B. Hormones, Vitamins, etc.

Other groups that can be included are, for example: hormones, steroids, androgens, estrogens, thyroid hormone, or vitamins, folic acid.

C. Polyalkylenes, Polysaccharides, etc.

20 Also, polyalkylene glycol can be included with the desired polynucleotides/polypeptides. In a preferred embodiment, the polyalkylene glycol is polyethylene glycol. In addition, mono-, di-, or polysaccharides can be included. In a preferred embodiment of this aspect, the polysaccharide is dextran or DEAE-dextran. Also, chitosan and poly(lactide-co-glycolide)

D. Lipids, and Liposomes

25 The desired polynucleotide/polypeptide can also be encapsulated in lipids or packaged in liposomes prior to delivery to the subject or to cells derived therefrom.

30 Lipid encapsulation is generally accomplished using liposomes which are able to stably bind or entrap and retain nucleic acid. The ratio of condensed polynucleotide to lipid preparation can vary but will generally be around 1:1 (mg DNA:micromoles lipid), or more of lipid. For a review of the use of liposomes as carriers for delivery of nucleic acids, see, Hug and Sleight (1991) *Biochim. Biophys. Acta.* 1097:1-17; Straubinger (1983) *Meth. Enzymol.* 101:512-527.

Liposomal preparations for use in the present invention include cationic (positively charged), anionic (negatively charged) and neutral preparations. Cationic liposomes have been shown to mediate intracellular delivery of plasmid DNA (Felgner (1987) *Proc. Natl. Acad. Sci. USA* 84:7413-7416); mRNA (Malone (1989) *Proc. Natl. Acad. Sci. USA* 86:6077-6081); and purified transcription factors (Debs (1990) *J. Biol. Chem.* 265:10189-10192), in functional form.

Cationic liposomes are readily available. For example, N[1-2,3-dioleoyloxy)propyl]-N,N,N-triethylammonium (DOTMA) liposomes are available under the trademark Lipofectin, from GIBCO BRL, Grand Island, NY. (See,

also, Felgner *supra*). Other commercially available liposomes include transfectace (DDAB/DOPE) and DOTAP/DOPE (Boehringer). Other cationic liposomes can be prepared from readily available materials using techniques well known in the art. See, e.g. Szoka (1978) *Proc. Natl. Acad. Sci. USA* 75:4194-4198; WO90/11092 for a description of the synthesis of DOTAP (1,2-bis(oleoyloxy)-3-(trimethylammonio)propane) liposomes.

Similarly, anionic and neutral liposomes are readily available, such as from Avanti Polar Lipids (Birmingham, AL), or can be easily prepared using readily available materials. Such materials include phosphatidyl choline, cholesterol, phosphatidyl ethanolamine, dioleoylphosphatidyl choline (DOPC), dioleoylphosphatidyl glycerol (DOPG), dioleoylphosphatidyl ethanolamine (DOPE), among others. These materials can also be mixed with the DOTMA and DOTAP starting materials in appropriate ratios. Methods for making liposomes using these materials are well known in the art.

The liposomes can comprise multilamellar vesicles (MLVs), small unilamellar vesicles (SUVs), or large unilamellar vesicles (LUVs). The various liposome-nucleic acid complexes are prepared using methods known in the art. See e.g. Straubinger (1983) *Meth. Immunol.* 101:512-527; Szoka (1978) *Proc. Natl. Acad. Sci. USA* 75:4194-4198; Papahadjopoulos (1975) *Biochim. Biophys. Acta* 394:483; Wilson (1979) *Cell* 17:77; Deamer & Bangham (1976) *Biochim. Biophys. Acta* 443:629; Ostro (1977) *Biochem. Biophys. Res. Commun.* 76:836; Fraley (1979) *Proc. Natl. Acad. Sci. USA* 76:3348; Enoch & Strittmatter (1979) *Proc. Natl. Acad. Sci. USA* 76:145; Fraley (1980) *J. Biol. Chem.* (1980) 255:10431; Szoka & Papahadjopoulos (1978) *Proc. Natl. Acad. Sci. USA* 75:145; and Schaefer-Ridder (1982) *Science* 215:166.

#### E. Lipoproteins

In addition, lipoproteins can be included with the polynucleotide/polypeptide to be delivered. Examples of lipoproteins to be utilized include: chylomicrons, HDL, IDL, LDL, and VLDL. Mutants, fragments, or fusions of these proteins can also be used. Also, modifications of naturally occurring lipoproteins can be used, such as acetylated LDL. These lipoproteins can target the delivery of polynucleotides to cells expressing lipoprotein receptors. Preferably, if lipoproteins are including with the polynucleotide to be delivered, no other targeting ligand is included in the composition.

Naturally occurring lipoproteins comprise a lipid and a protein portion. The protein portion are known as apoproteins. At the present, apoproteins A, B, C, D, and E have been isolated and identified. At least two of these contain several proteins, designated by Roman numerals, AI, AII, AIV; CI, CII, CIII.

A lipoprotein can comprise more than one apoprotein. For example, naturally occurring chylomicrons comprises of A, B, C, & E, over time these lipoproteins lose A and acquire C and E apoproteins. VLDL comprises A, B, C, & E apoproteins, LDL comprises apoprotein B; HDL comprises apoproteins A, C, & E.

The amino acid of these apoproteins are known and are described in, for example, Breslow (1985) *Annu Rev. Biochem* 54:699; Law (1986) *Adv. Exp Med. Biol.* 151:162; Chen (1986) *J Biol Chem* 261:12918; Kane (1980) *Proc Natl Acad Sci USA* 77:2465; and Utermann (1984) *Hum Genet* 65:232.

Lipoproteins contain a variety of lipids including, triglycerides, cholesterol (free and esters), and phospholipids. The composition of the lipids varies in naturally occurring lipoproteins. For example, chylomicrons comprise mainly triglycerides. A more detailed description of the lipid content of naturally occurring lipoproteins can be found, for example, in *Meth. Enzymol.* 128 (1986). The composition of the lipids are chosen to aid in

conformation of the apoprotein for receptor binding activity. The composition of lipids can also be chosen to facilitate hydrophobic interaction and association with the polynucleotide binding molecule.

Naturally occurring lipoproteins can be isolated from serum by ultracentrifugation, for instance. Such methods are described in *Meth. Enzymol. (supra)*; Pitas (1980) *J. Biochem.* 255:5454-5460 and Mahey (1979) *J Clin. Invest* 64:743-750. Lipoproteins can also be produced by *in vitro* or recombinant methods by expression of the apoprotein genes in a desired host cell. See, for example, Atkinson (1986) *Annu Rev Biophys Chem* 15:403 and Radding (1958) *Biochim Biophys Acta* 30: 443. Lipoproteins can also be purchased from commercial suppliers, such as Biomedical Technologies, Inc., Stoughton, Massachusetts, USA. Further description of lipoproteins can be found in Zuckermann *et al.* PCT/US97/14465.

#### 10 F. Polycationic Agents

Polycationic agents can be included, with or without lipoprotein, in a composition with the desired polynucleotide/polypeptide to be delivered.

Polycationic agents, typically, exhibit a net positive charge at physiological relevant pH and are capable of neutralizing the electrical charge of nucleic acids to facilitate delivery to a desired location. These agents have both *in vitro*, *ex vivo*, and *in vivo* applications. Polycationic agents can be used to deliver nucleic acids to a living subject either intramuscularly, subcutaneously, etc.

The following are examples of useful polypeptides as polycationic agents: polylysine, polyarginine, polyornithine, and protamine. Other examples include histones, protamines, human serum albumin, DNA binding proteins, non-histone chromosomal proteins, coat proteins from DNA viruses, such as (X174, transcriptional factors also contain domains that bind DNA and therefore may be useful as nucleic acid condensing agents. Briefly, transcriptional factors such as C/EBP, c-jun, c-fos, AP-1, AP-2, AP-3, CPF, Prot-1, Sp-1, Oct-1, Oct-2, CREP, and TFIID contain basic domains that bind DNA sequences.

Organic polycationic agents include: spermine, spermidine, and putrescine.

The dimensions and of the physical properties of a polycationic agent can be extrapolated from the list above, to construct other polypeptide polycationic agents or to produce synthetic polycationic agents.

Synthetic polycationic agents which are useful include, for example, DEAE-dextran, polybrene. Lipofectin™, and lipofectAMINE™ are monomers that form polycationic complexes when combined with polynucleotides/polypeptides.

#### Nucleic Acid Hybridisation

"Hybridization" refers to the association of two nucleic acid sequences to one another by hydrogen bonding. Typically, one sequence will be fixed to a solid support and the other will be free in solution. Then, the two sequences will be placed in contact with one another under conditions that favor hydrogen bonding. Factors that affect this bonding include: the type and volume of solvent; reaction temperature; time of hybridization; agitation; agents to block the non-specific attachment of the liquid phase sequence to the solid support (Denhardt's reagent or BLOTTO); concentration of the sequences; use of compounds to increase the rate of association of sequences (dextran sulfate or polyethylene glycol); and the stringency of the washing conditions following hybridization. See Sambrook *et al.* [*supra*] vol.2, chapt.9, pp.9.47 to 9.57.

"Stringency" refers to conditions in a hybridization reaction that favor association of very similar sequences over sequences that differ. For example, the combination of temperature and salt concentration should be chosen that is approximately 120 to 200°C below the calculated  $T_m$  of the hybrid under study. The temperature and salt conditions can often be determined empirically in preliminary experiments in which samples of genomic DNA immobilized on filters are hybridized to the sequence of interest and then washed under conditions of different stringencies. See Sambrook *et al.* at page 9.50.

Variables to consider when performing, for example, a Southern blot are (1) the complexity of the DNA being blotted and (2) the homology between the probe and the sequences being detected. The total amount of the fragment(s) to be studied can vary a magnitude of 10, from 0.1 to 1 µg for a plasmid or phage digest to  $10^{-9}$  to  $10^{-8}$  g for a single copy gene in a highly complex eukaryotic genome. For lower complexity polynucleotides, substantially shorter blotting, hybridization, and exposure times, a smaller amount of starting polynucleotides, and lower specific activity of probes can be used. For example, a single-copy yeast gene can be detected with an exposure time of only 1 hour starting with 1 µg of yeast DNA, blotting for two hours, and hybridizing for 4-8 hours with a probe of  $10^8$  cpm/µg. For a single-copy mammalian gene a conservative approach would start with 10 µg of DNA, blot overnight, and hybridize overnight in the presence of 10% dextran sulfate using a probe of greater than  $10^8$  cpm/µg, resulting in an exposure time of ~24 hours.

Several factors can affect the melting temperature ( $T_m$ ) of a DNA-DNA hybrid between the probe and the fragment of interest, and consequently, the appropriate conditions for hybridization and washing. In many cases the probe is not 100% homologous to the fragment. Other commonly encountered variables include the length and total G+C content of the hybridizing sequences and the ionic strength and formamide content of the hybridization buffer. The effects of all of these factors can be approximated by a single equation:

$$T_m = 81 + 16.6(\log_{10} C_i) + 0.4[\%(G + C)] - 0.6(\% \text{ formamide}) - 600/n - 1.5(\% \text{ mismatch}).$$

where  $C_i$  is the salt concentration (monovalent ions) and  $n$  is the length of the hybrid in base pairs (slightly modified from Meinkoth & Wahl (1984) *Anal. Biochem.* 138: 267-284).

In designing a hybridization experiment, some factors affecting nucleic acid hybridization can be conveniently altered. The temperature of the hybridization and washes and the salt concentration during the washes are the simplest to adjust. As the temperature of the hybridization increases (*ie.* stringency), it becomes less likely for hybridization to occur between strands that are nonhomologous, and as a result, background decreases. If the radiolabeled probe is not completely homologous with the immobilized fragment (as is frequently the case in gene family and interspecies hybridization experiments), the hybridization temperature must be reduced, and background will increase. The temperature of the washes affects the intensity of the hybridizing band and the degree of background in a similar manner. The stringency of the washes is also increased with decreasing salt concentrations.

In general, convenient hybridization temperatures in the presence of 50% formamide are 42°C for a probe with 95% to 100% homologous to the target fragment, 37°C for 90% to 95% homology, and 32°C for 85% to 90% homology. For lower homologies, formamide content should be lowered and temperature adjusted accordingly, using the equation above. If the homology between the probe and the target fragment are not known, the simplest approach is to start with both hybridization and wash conditions which are nonstringent. If non-specific bands or high background are observed after autoradiography, the filter can be washed at high stringency and

reexposed. If the time required for exposure makes this approach impractical, several hybridization and/or washing stringencies should be tested in parallel.

#### Nucleic Acid Probe Assays

Methods such as PCR, branched DNA probe assays, or blotting techniques utilizing nucleic acid probes according to the invention can determine the presence of cDNA or mRNA. A probe is said to "hybridize" with a sequence of the invention if it can form a duplex or double stranded complex, which is stable enough to be detected.

The nucleic acid probes will hybridize to the Chlamydial nucleotide sequences of the invention (including both sense and antisense strands). Though many different nucleotide sequences will encode the amino acid sequence, the native Chlamydial sequence is preferred because it is the actual sequence present in cells. mRNA represents a coding sequence and so a probe should be complementary to the coding sequence; single-stranded cDNA is complementary to mRNA, and so a cDNA probe should be complementary to the non-coding sequence.

The probe sequence need not be identical to the Chlamydial sequence (or its complement) — some variation in the sequence and length can lead to increased assay sensitivity if the nucleic acid probe can form a duplex with target nucleotides, which can be detected. Also, the nucleic acid probe can include additional nucleotides to stabilize the formed duplex. Additional Chlamydial sequence may also be helpful as a label to detect the formed duplex. For example, a non-complementary nucleotide sequence may be attached to the 5' end of the probe, with the remainder of the probe sequence being complementary to a Chlamydial sequence. Alternatively, non-complementary bases or longer sequences can be interspersed into the probe, provided that the probe sequence has sufficient complementarity with the a Chlamydial sequence in order to hybridize therewith and thereby form a duplex which can be detected.

The exact length and sequence of the probe will depend on the hybridization conditions, such as temperature, salt condition and the like. For example, for diagnostic applications, depending on the complexity of the analyte sequence, the nucleic acid probe typically contains at least 10-20 nucleotides, preferably 15-25, and more preferably  $\geq 30$  nucleotides, although it may be shorter than this. Short primers generally require cooler temperatures to form sufficiently stable hybrid complexes with the template.

Probes may be produced by synthetic procedures, such as the triester method of Matteucci *et al.* [*J. Am. Chem. Soc.* (1981) 103:3185], or according to Urdea *et al.* [*Proc. Natl. Acad. Sci. USA* (1983) 80: 7461], or using commercially available automated oligonucleotide synthesizers.

The chemical nature of the probe can be selected according to preference. For certain applications, DNA or RNA are appropriate. For other applications, modifications may be incorporated *e.g.* backbone modifications, such as phosphorothioates or methylphosphonates, can be used to increase *in vivo* half-life, alter RNA affinity, increase nuclease resistance *etc.* [*e.g.* see Agrawal & Iyer (1995) *Curr Opin Biotechnol* 6:12-19; Agrawal (1996) *TIBTECH* 14:376-387]; analogues such as peptide nucleic acids may also be used [*e.g.* see Corey (1997) *TIBTECH* 15:224-229; Buchardt *et al.* (1993) *TIBTECH* 11:384-386].

Alternatively, the polymerase chain reaction (PCR) is another well-known means for detecting small amounts of target nucleic acids. The assay is described in: Mullis *et al.* [*Meth. Enzymol.* (1987) 155: 335-350]; US patents 4,683,195 & 4,683,202. Two 'primers' hybridize with the target nucleic acids and are used to prime the reaction. The primers can comprise sequence that does not hybridize to the sequence of the amplification target (or its

complement) to aid with duplex stability or, for example, to incorporate a convenient restriction site. Typically, such sequence will flank the desired Chlamydial sequence.

A thermostable polymerase creates copies of target nucleic acids from the primers using the original target nucleic acids as a template. After a threshold amount of target nucleic acids are generated by the polymerase, they can be detected by more traditional methods, such as Southern blots. When using the Southern blot method, the labelled probe will hybridize to the Chlamydial sequence (or its complement).

Also, mRNA or cDNA can be detected by traditional blotting techniques described in Sambrook *et al* [*supra*]. mRNA, or cDNA generated from mRNA using a polymerase enzyme, can be purified and separated using gel electrophoresis. The nucleic acids on the gel are then blotted onto a solid support, such as nitrocellulose. The solid support is exposed to a labelled probe and then washed to remove any unhybridized probe. Next, the duplexes containing the labeled probe are detected. Typically, the probe is labelled with a radioactive moiety.

### BRIEF DESCRIPTION OF THE DRAWINGS

Figures 1-189 show data pertaining to examples 1-189.

Figure 190 shows a representative 2D gel of proteins in elementary bodies.

Figure 191 shows an alignment of sequences in five (six) proteins of the invention.

### EXAMPLES

The examples indicate *C.pneumoniae* proteins, together with evidence to support the view that the proteins are useful antigens for vaccine production and development or for diagnostic purposes. This evidence takes the form of:

- Computer prediction based on sequence information from CWL029 strain (*e.g.* using the PSORT algorithm available from [www.psort.nibb.ac.jp](http://www.psort.nibb.ac.jp)).
- Data on recombinant expression and purification of the proteins cloned from IOL207 strain.
- Western blots to demonstrate immunoreactivity in serum (typically a blot of an EB extract of *C.pneumoniae* strain FB/96 stained with mouse antiserum against the recombinant protein).
- FACS analysis of *C.pneumoniae* bacteria or purified EBs to confirm accessibility of the antigen to the immune system (see also table III).
- An indication if the protein was identified by MALDI-TOF from a 2D gel electrophoresis map of proteins from purified elementary bodies from strain FB/96. This confirms that the protein is expressed *in vivo* (see also table V).

Various tests can be used to assess the *in vivo* immunogenicity of the proteins identified in the examples. For example, the proteins can be expressed recombinantly and used to screen patient sera by immunoblot. A positive reaction between the protein and patient serum indicates that the patient has previously mounted an immune response to the protein in question *ie.* the protein is an immunogen. This method can also be used to identify immunodominant proteins.

The recombinant protein can also be conveniently used to prepare antibodies *e.g.* in a mouse. These can be used for direct confirmation that a protein is located on the cell-surface. Labelled antibody (*e.g.* fluorescent labelling for FACS) can be incubated with intact bacteria and the presence of label on the bacterial surface confirms the location of the protein.

- 5 In particular, the following methods (A) to (O) were used to express, purify and biochemically characterise the proteins of the invention:

#### CLONING OF CPN ORFs FOR EXPRESSION IN *E. COLI*

ORFs of *Chlamydia pneumoniae* (Cpn) were cloned in such a way as to potentially obtain three different kind of proteins:

- 10 a) proteins having an hexa-histidine tag at the C-terminus (cpn-His)  
 b) proteins having a GST fusion partner at the N-terminus (Gst-cpn)  
 c) proteins having both hexa-histidine tag at the C-terminus and GST at the N-terminus (GST/His fusion; NH<sub>2</sub>-GST-cpn-(His)<sub>6</sub>-COOH)

- The type a) proteins were obtained upon cloning in the pET21b+ (Novagen). The type b) and c) proteins were obtained upon cloning in modified pGEX-KG vectors [Guan & Dixon (1991) *Anal. Biochem.* 192:262]. For instance pGEX-KG was modified to obtain pGEX-NN, then by modifying pGEX-NN to obtain pGEX-NNH. The Gst-cpn and Gst-cpn-His proteins were obtained in pGEX-NN and pGEX-NNH respectively.

- 20 The modified versions of pGEX-KG vector were made with the aim of allowing the cloning of single amplification products in all three vectors after only one double restriction enzyme digestion and to minimise the presence of extraneous amino acids in the final recombinant proteins.

#### (A) Construction of pGEX-NN and pGEX-NNH expression vectors

- Two couples of complementary oligodeoxyribonucleotides were synthesised using the DNA synthesiser ABI394 (Perkin Elmer) and the reagents from Cruachem (Glasgow, Scotland). Equimolar amounts of the oligo pairs (50 ng each oligo) were annealed in T4 DNA ligase buffer (New England Biolabs) for 10 min in a final volume of 50µl and then were left to cool slowly at room temperature. With the described procedure the following DNA linkers were obtained:

##### gexNN linker:

- 30 NdeI NheI XmaI EcoRI NcoI SalI XhoI SacI NotI  
 GATCCCATATGGCTAGCCCGGGGAATTCGTCATGGAGTGAGTCGACTGACTCGAGTGATCGAGCTCCTGAGCGGCCGCATGAA  
 GGTATACCGATCGGGCCCCCTAAGCAGGTACCTCACTCAGCTGACTGAGCTCACTAGCTCGAGGACTCGCCGGCGTACTTTCTGA

##### gexNNH linker:

- 35 HindIII NotI XhoI --Hexa-Histidine--  
 TCGACAAGCTTGCGGCCGCACTCGAGCATCACCATCACCATCACTGAT  
 GTTCGAACGCCGGCGTGAGCACGTAGAGGTAGTGGTAGTGACTATCGA

The plasmid pGEX-KG was digested with BamHI and HindIII and 100 ng were ligated overnight at 16 °C to the linker gexNN with a molar ratio of 3:1 linker/plasmid using 200 units of T4 DNA ligase

(New england Biolabs). After transformation of the ligation product in *E. coli* DH5, a clone containing the pGEX-NN plasmid, having the correct linker, was selected by means of restriction enzyme analysis and DNA sequencing.

The new plasmid pGEX-NN was digested with SalI and HindIII and ligated to the linker gexNNH.

- 5 After transformation of the ligation product in *E. coli* DH5, a clone containing the pGEX-NNH plasmid, having the correct linker, was selected by means of restriction enzyme analysis and DNA sequencing.

### (B) Chromosomal DNA preparation

- 10 The chromosomal DNA of elementary bodies (EB) of *C.pneumoniae* strain 10L-207 was prepared by adding 1.5 ml of lysis buffer (10 mM Tris-HCl, 150 mM NaCl, 2 mM EDTA, 0,6 % SDS, 100 µg/ml Proteinase K, pH 8) to 450 µl EB suspension (400.000/µl) and incubating overnight at 37 °C. After sequential extraction with phenol, phenol-chloroform, and chloroform, the DNA was precipitated with 0,3 M sodium acetate, pH 5,2 and 2 volumes of absolute ethanol. The DNA pellet was washed with 70 % ethanol. After solubilization with distilled water and treatment with 20 µg/ml RNase A
- 15 for 1 hour at RT, the DNA was extracted again with phenol-chloroform, alcohol precipitated and suspended with 300 µl 1 mM Tris-HCl pH 8,5. The DNA concentration was evaluated by measuring OD<sub>260</sub> of the sample.

### (C) Oligonucleotide design

- 20 Synthetic oligonucleotide primers were designed on the basis of the coding sequence of each ORF using the sequence of *C.pneumoniae* strain CWL029. Any predicted signal peptide were omitted, by deducing the 5' end amplification primer sequence immediately downstream from the predicted leader sequence. For most ORFs, the 5' tail of the primers (table I) included only one restriction enzyme recognition site (NdeI, or NheI, or SpeI depending on the gene's own restriction pattern); the 3' primer tails (tableI) included a XhoI or a NotI or a HindIII restriction site.

5' tails		3' tails	
NdeI	5' GTGCGTCATATG 3'	XhoI	5' GCGTCTCGAG 3'
NheI	5' GTGCGTGCTAGC 3'	NotI	5' ACTCGCTAGCGGCCGC 3'
SpeI	5' GTGCGTACTAGT 3'	HindIII	5' GCGTAAGCTT 3'

25 **Table I.** Oligonucleotide tails of the primers used to amplify Cpn genes.

- As well as containing the restriction enzyme recognition sequences, the primers included nucleotides which hybridized to the sequence to be amplified. The number of hybridizing nucleotides depended on the melting temperature of the primers which was determined as described [(Breslauer *et al.* (1986) *PNAS USA* 83:3746-50]. The average melting temperature of the selected oligos was 50-55°C
- 30 for the hybridizing region alone and 65-75°C for the whole oligos. Table II shows the forward and reverse primers used for each amplification.



**(D) Amplification**

The standard PCR protocol was as follow: 50 ng genomic DNA were used as template in the presence of 0,2  $\mu$ M each primer, 200  $\mu$ M each dNTP, 1,5 mM  $MgCl_2$ , 1x PCR buffer minus Mg (Gibco-BRL), and 2 units of Taq DNA polymerase (Platinum Taq, Gibco-BRL) in a final volume of 100  $\mu$ l. Each sample underwent a double-step amplification: the first 5 cycles were performed using as the hybridizing temperature the one of the oligos excluding the restriction enzyme tail, followed by 25 cycles performed according to the hybridization temperature of the whole lenght primers. The standard cycles were as follow:

denaturation : 94 °C, 2 min

denaturation: 94 °C, 30 seconds	}	5 cycles
hybridization: 51 °C, 50 seconds		
elongation: 72 °C, 1 min or 2 min and 40 sec		

denaturation: 94 °C, 30 seconds	}	25 cycles
hybridization: 70 °C, 50 seconds		
elongation: 72 °C, 1 min or 2 min and 40 sec		

72 °C, 7 min

4 °C

The elongation time was 1 min for ORFs shorter than 2000 bp, and 2 min and 40 seconds for ORFs longer than 2000 bp. The amplifications were performed using a Gene Amp PCR system 9600 (Perkin Elmer).

To check the amplification results, 4  $\mu$ l of each PCR product was loaded onto 1-1.5 agarose gel and the size of amplified fragments compared with DNA molecular weight standards (DNA markers III or IX, Roche). The PCR products were loaded on agarose gel and after electrophoresis the right size bands were excised from the gel. The DNA was purified from the agarose using the Gel Extraction Kit (Qiagen) following the instruction of the manufacturer. The final elution volume of the DNA was 50  $\mu$ l TE (10 mM Tris-HCl, 1 mM EDTA, pH 8). One  $\mu$ l of each purified DNA was loaded onto agarose gel to evaluate the yield.

**(E) Digestion of PCR fragments**

One-two  $\mu$ g of purified PCR product were double digested overnight at 37 °C with the appropriate restriction enzymes (60 units of each enzyme) using the appropriate restriction buffer in 100  $\mu$ l final volume. The restriction enzymes and the digestion buffers were from New England Biolabs. After

purification of the digested DNA (PCR purification Kit, Qiagen) and elution with 30 µl TE, 1 µl was subjected to agarose gel electrophoresis to evaluate the yield in comparison to titrated molecular weight standards (DNA markers III or IX, Roche).

**(F) Digestion of the cloning vectors (pET21b+, pGEX-NN, and pGEX-NNH)**

- 5 10 µg of plasmid was double digested with 100 units of each restriction enzyme in 400 µl reaction volume in the presence of appropriate buffer by overnight incubation at 37 °C. After electrophoresis on a 1% agarose gel, the band corresponding to the digested vector was purified from the gel using the Qiagen Qiaex II Gel Extraction Kit and the DNA was eluted with 50 µl TE. The DNA concentration was evaluated by measuring OD<sub>260</sub> of the sample.

10 **(G) Cloning**

75ng of the appropriately digested and purified vectors and the digested and purified fragments corresponding to each ORF, were ligated in final volumes of 10-20 µl with a molar ratio of 1:1 fragment/vector, using 400 units T4 DNA ligase (New England Biolabs) in the presence of the buffer supplied by the manufacturer. The reactions were incubated overnight at 16 °C.

- 15 Transformation in *E. coli* DH5 competent cells was performed as follow: the ligation reaction was mixed with 200 µl of competent DH5 cells and incubated on ice for 30 min and then at 42 °C for 90 seconds. After cooling on ice, 0.8 ml LB was added and the cells were incubated for 45 min at 37 °C under shaking. 100 and 900 µl of cell suspensions were plated on separate plates of agar LB 100 µg/ml Ampicillin and the plates were incubated overnight at 37 °C. The screening of the
- 20 transformants was done by growing randomly chosen clones in 6 ml LB 100 µg/ml Ampicillin, by extracting the DNA using the Qiagen Qiaprep Spin Miniprep Kit following the manufacturer instructions, and by digesting 2 µl of plasmid miniprep with the restriction enzymes specific for the restriction cloning sites. After agarose gel electrophoresis of the digested plasmid mini-preparations, positive clones were chosen on the basis of the correct size of the restriction fragments,
- 25 as evaluated by comparison with appropriate molecular weight markers (DNA markers III or IX, Roche).

**(H) Expression**

- 1 µl of each right plasmid mini-preparation was transformed in 200 µl of competent *E. coli* strain suitable for expression of the recombinant protein. All pET21b+ recombinant plasmids were
- 30 transformed in BL21 DE3 (Novagen) *E. coli* cells, whilst all pGEX-NN and all pGEX-NNH recombinant plasmids were transformed in BL21 cells (Novagen). After plating transformation mixtures on LB/Amp agar plates and incubation overnight at 37 °C, single colonies were inoculated in 3 ml LB 100 µg/ml Ampicillin and grown at 37 °C overnight. 70 µl of the overnight culture was inoculated in 2 ml LB/Amp and grown at 37 °C until OD<sub>600</sub> of the pET clones reached the 0,4-0,8
- 35 value or until OD<sub>600</sub> of the pGEX clones reached the 0,8-1 value. Protein expression was then

induced by adding IPTG (Isopropil  $\beta$ -D thio-galacto-piranoside) to the mini-cultures. pET clones were induced using 1 mM IPTG, whilst pGEX clones were induced using 0.2 mM IPTG. After 3 hours incubation at 37 °C the final OD<sub>600</sub> was checked and the cultures were cooled on ice. After centrifugation of 0.5 ml culture, the cell pellet was suspended in 50  $\mu$ l of protein Loading Sample Buffer (60 mM TRIS-HCl pH 6.8, 5% w/v SDS, 10% v/v glycerin, 0.1% w/v Bromophenol Blue, 100 mM DTT) and incubated at 100 °C for 5 min. A volume of boiled sample corresponding to 0.1 OD<sub>600</sub> culture was analysed by SDS-PAGE and Coomassie Blue staining to verify the presence of induced protein band.

#### PURIFICATION OF THE RECOMBINANT PROTEINS

- Single colonies were inoculated in 25 ml LB 100  $\mu$ g/ml Ampicillin and grown at 37 °C overnight. The overnight culture was inoculated in 500 ml LB/Amp and grown under shaking at 25 °C until OD<sub>600</sub> 0,4-0,8 value for the pET clones, or until OD<sub>600</sub> 0,8-1 value for the pGEX clones. Protein expression was then induced by adding IPTG to the cultures. pET clones were induced using 1 mM IPTG, whilst pGEX clones were induced using 0.2 mM IPTG. After 4 hours incubation at 25 °C the final OD<sub>600</sub> was checked and the cultures were cooled on ice. After centrifugation at 6000 rpm (JA10 rotor, Beckman), the cell pellet was processed for purification or frozen at -20 °C.

##### (I) Procedure for the purification of soluble His-tagged proteins from *E.coli*

1. Transfer the pellets from -20°C to ice bath and reconstitute with 10 ml 50 mM NaHPO<sub>4</sub> buffer, 300 mM NaCl, pH 8,0, pass in 40-50 ml centrifugation tubes and break the cells as per the following outline:
2. Break the pellets in the French Press performing three passages with in-line washing.
3. Centrifuge at about 30-40000 x g per 15-20 min. If possible use rotor JA 25.50 (21000 rpm, 15 min.) or JA-20 (18000 rpm, 15 min.)
4. Equilibrate the Poly-Prep columns with 1 ml Fast Flow Chelating Sepharose resin with 50 mM phosphate buffer, 300 mM NaCl, pH 8,0.
5. Store the centrifugation pellet at -20°C, and load the supernatant in the columns.
6. Collect the flow through.
7. Wash the columns with 10 ml (2 ml + 2 ml + 4 ml) 50 mM phosphate buffer, 300 mM NaCl, pH 8,0.
8. Wash again with 10 ml 20 mM imidazole buffer, 50 mM phosphate, 300 mM NaCl, pH 8,0.
9. Elute the proteins bound to the columns with 4,5 ml (1,5 ml + 1,5 ml + 1,5 ml) 250 mM imidazole buffer, 50 mM phosphate, 300 mM NaCl, pH 8,0 and collect the 3 corresponding fractions of ~1,5 ml each. Add to each tube 15  $\mu$ l DTT 200 mM (final concentration 2 mM)

10. Measure the protein concentration of the first two fractions with the Bradford method, collect a 10 µg aliquot of proteins from each sample and analyse by SDS-PAGE. (N.B.: should the sample be too diluted, load 21 µl + 7 µl loading buffer).
11. Store the collected fractions at +4°C while waiting for the results of the SDS-PAGE analysis.
- 5 12. For immunisation prepare 4-5 aliquots of 100 µg each in 0,5 ml in 40% glycerol. The dilution buffer is the above elution buffer, plus 2 mM DTT. Store the aliquots at -20°C until immunisation.

#### **(J) Purification of His-tagged proteins from Inclusion bodies**

Purifications were carried out essentially according the following protocol:

- 10 1. Bacteria are collected from 500 ml cultures by centrifugation. If required store bacterial pellets at -20°C. For extraction, resuspend each bacterial pellet in 10 ml 50 mM TRIS-HCl buffer, pH 8,5 on an ice bath.
2. Disrupt the resuspended bacteria with a French Press, performing two passages.
3. Centrifuge at 35000 x g for 15 min and collect the pellets. Use a Beckman rotor JA 25.50 (21000  
15 rpm, 15 min.) or JA-20 (18000 rpm, 15 min.).
4. Dissolve the centrifugation pellets with 50 mM TRIS-HCl, 1 mM TCEP {Tris(2-carboxyethyl)-phosphine hydrochloride, Pierce} , 6M guanidium chloride, pH 8,5. Stir for ~ 10 min. with a magnetic bar.
5. Centrifuge as described above, and collect the supernatant..
- 20 6. Prepare an adequate number of Poly-Prep (Bio-Rad) columns containing 1 ml of Fast Flow Chelating Sepharose (Pharmacia) saturated with Nichel according to manufacturer recommendations.. Wash the columns twice with 5 ml of H<sub>2</sub>O and equilibrate with 50 mM TRIS-HCl, 1 mM TCEP, 6M guanidinium chloride, pH 8,5.
7. Load the supernatants from step 5 onto the columns, and wash with 5 ml of 50 mM TRIS-Hcl  
25 buffer, 1 mM TCEP, 6M urea, pH 8,5
8. Wash the columns with 10 ml of 20 mM imidazole, 50 mM TRIS-HCl , 6M urea, 1 mM TCEP, pH 8,5. Collect and set aside the first 5 ml for possible further controls.
9. Elute the proteins bound to the columns with 4,5 ml of a buffer containing 250 mM imidazole, 50 mM TRIS-HCl, 6M urea, 1 mM TCEP, pH 8,5. Add the elution buffer in three 1,5 ml aliquots,  
30 and collect the corresponding 3 fractions. Add to each fraction 15 µl DTT (final concentration 2 mM) .
10. Measure eluted protein concentration with the Bradford method, and analyze aliquots of ca 10 µg of protein by SDS-PAGE.
11. Store proteins at -20°C in 40% (v/v) glycerol, 50 mM TRIS-HCl, 2M urea, 0.5 M arginine, 2 mM  
35 DTT, 0.3 mM TCEP, 83.3 mM imidazole, pH 8,5

**(K) Procedure for the purification of GST-fusion proteins from *E.coli***

1. Transfer the bacterial pellets from  $-20^{\circ}\text{C}$  to an ice bath and resuspend with 7,5 ml PBS, pH 7,4 to which a mixture of protease inhibitors (CØMPLETE™ - Boehringer Mannheim, 1 tablet every 25 ml of buffer) has been added. Transfer to 40-50 ml centrifugation tubes and sonicate according to the following procedure:
  - a) Position the probe at about 0,5 cm from the bottom of the tube
  - b) Block the tube with the clamp
  - c) Dip the tube in an ice bath
  - d) Set the sonicator as follows: Timer  $\rightarrow$  Hold, Duty Cycle  $\rightarrow$  55, Out. Control  $\rightarrow$  6.
  - e) perform 5 cycles of 10 impulses at a time lapse of 1 minute (i.e. one cycle = 10 impulses +  $\sim 45''$  hold; b. 10 impulses +  $\sim 45''$  hold; c. 10 impulses +  $\sim 45''$  hold; d. 10 impulses +  $\sim 45''$  hold; e. 10 impulses +  $\sim 45''$  hold)
2. Centrifuge at about  $30-40000 \times g$  for 15-20 min. E.g.: use rotor Beckman JA 25.50 at 21000 rpm, for 15 min.
3. Store the centrifugation pellets at  $-20^{\circ}\text{C}$ , and load the supernatants on the chromatography columns, as follows
4. Equilibrate the Poly-Prep (Bio-Rad) columns with 0,5 ml ( $\cong 1$  ml suspension) of Glutathione-Sepharose 4B resin, wash with 2 ml (1 + 1)  $\text{H}_2\text{O}$ , and then with 10 ml (2 + 4 + 4) PBS, pH 7,4.
5. Load the supernatants on the columns and discard the flow through.
6. Wash the columns with 10 ml (2 + 4 + 4) PBS, pH 7,4.
7. Elute the proteins bound to the columns with 4,5 ml of 50 mM TRIS buffer, 10 mM reduced glutathione, pH 8,0, adding 1,5 ml + 1,5 ml + 1,5 ml and collecting the respective 3 fractions of  $\sim 1,5$  ml each.
8. Measure the protein concentration of the first two fractions with the Bradford method, analyse a 10  $\mu\text{g}$  aliquot of proteins from each sample by SDS-PAGE. (N.B.: if the sample is too diluted load 21  $\mu\text{l}$  (+ 7  $\mu\text{l}$  loading buffer).
9. Store the collected fractions at  $+4^{\circ}\text{C}$  while waiting for the results of the SDS-PAGE analysis.
10. For each protein destined to the immunisation prepare 4-5 aliquots of 100  $\mu\text{g}$  each in 0,5 ml of 40% glycerol. The dilution buffer is 50 mM TRIS.HCl, 2 mM DTT, pH 8,0. Store the aliquots at  $-20^{\circ}\text{C}$  until immunisation..

**SEROLOGY****(L) Protocol of immunization**

1. Groups of four CD1 female mice aged between 6 and 7 weeks were immunized with 20  $\mu\text{g}$  of recombinant protein resuspended in 100  $\mu\text{l}$ .

2. Four mice for each group received 3 doses with a 14 days interval schedule.
3. Immunization was performed through intra-peritoneal injection of the protein with an equal volume of Complete Freund's Adjuvant (CFA) for the first dose and Incomplete Freund's Adjuvant (IFA) for the following two doses.
- 5 4. Sera were collected before each immunization. Mice were sacrificed 14 days after the third immunization and the collected sera were pooled and stored at -20°C.

**(M) Western blot analysis of Cpn elementary body proteins with mouse sera**

- Aliquots of elementary bodies containing approximately 4 µg of proteins, mixed with SDS loading buffer (1x: 60 mM TRIS-HCl pH 6.8, 5% w/v SDS, 10% v/v glycerin, 0.1% Bromophenol Blue, 100 mM DTT) and boiled 5 minutes at 95° C, were loaded on a 12% SDS-PAGE gel. The gel was run using a SDS-PAGE running buffer containing 250 mM TRIS, 2.5 mM Glycine and 0.1 %SDS. The gel was electroblotted onto nitrocellulose membrane at 200 mA for 30 minutes. The membrane was blocked for 30 minutes with PBS, 3% skimmed milk powder and incubated O/N at 4° C with the appropriate dilution (1/100) of the sera. After washing twice with PBS + 0.1% Tween (Sigma) the membrane was incubated for 2 hours with peroxidase-conjugated secondary anti-mouse antibody (Sigma) diluted 1:3000. The nitrocellulose was washed twice for 10 minutes with PBS + 0.1% Tween-20 and once with PBS and thereafter developed by Opti-4CN Substrate Kit (Biorad).

Lanes shown in Western blots are: (P) = pre-immune control serum; (I) = immune serum.

**(N) FACS analysis of *Chlamydia pneumoniae* elementary bodies with mouse sera**

- 20 1.  $2 \times 10^5$  Elementary Bodies (EB)/well were washed with 200 µl of PBS-0.1%BSA in a 96 wells U bottom plate and centrifuged for 10 min. at 1200rpm, at 4°C.
2. The supernatant was discarded and the E.B. resuspended in 10 µl of PBS-0.1%BSA.
3. 10µl mouse sera diluted in PBS-0.1%BSA were added to the E.B. suspension to a final dilution of 1:400, and incubated on ice for 30 min.
- 25 4. EB were washed by adding 180µl PBS-0.1%BSA and centrifuged for 10min. at 1200rpm, 4°C.
5. The supernatant was discarded and the E.B. resuspended in 10 l of PBS-0.1%BSA.
6. 10µl of a goat anti-mouse IgG, F(ab')<sub>2</sub> fragment specific-R-Phycoerythrin-conjugated (Jackson ImmunoResearch Laboratories Inc., cat.N°115-116-072) was added to the EB suspension to a final dilution of 1:100, and incubated on ice for 30 min. in the dark.
- 30 7. EB were washed by adding 180µl PBS-0.1%BSA and centrifuged for 10min. at 1200rpm, 4°C.
8. The supernatant was discarded and the E.B. resuspended in 150 µl of PBS-0.1%BSA.
9. E.B. suspension was passed through a cytometric chamber of a FACS Calibur (Becton Dickinson, Mountain View, CA USA) and 10.000 events were acquired.

10. Data were analysed using Cell Quest Software (Becton Dickinson, Mountain View, CA USA) by drawing a morphological dot plot (using forward and side scatter parameters) on E.B. signals. An histogram plot was then created on FL2 intensity of fluorescence log scale recalling the morphological region of EB.

- 5 NB: the results of FACS depend not only on the extent of accessibility of the native antigens but also on the quality of the antibodies elicited by the recombinant antigens, which may have structures with a variable degree of correct folding as compared with the native protein structures. Therefore, even if a FACS assay appears negative this does not necessarily mean that the protein is not abundant or accessible on the surface. PorB antigen, for instance, gave negative results in FACS but is a surface-exposed neutralising antigen [Kubo & Stephens (2000) *Mol. Microbiol.* 38:772-780].

#### (O) Mass Spectrometry analysis of two-dimensional electrophoretic protein maps

- Gradient purified EBs from strain FB/96 were solubilized at a final concentration of 5.5mg/ml with immobiline rehydration buffer (7M urea, 2M thiourea, 2% (w/v) CHAPS, 2% (w/v) ASB 14 [Chevallet *et al.* (1998) *Electrophor.* 19:1901-9], 2% (v/v) C.A 3-10NL (Amersham Pharmacia Biotech), 2 mM tributyl phosphine, 65 mM DTT). Samples (250µg protein) were adsorbed overnight on Immobiline DryStrips (7 cm, pH 3-10 non linear). Electrophocusing was performed in a IPhor Isoelectric Focusing Unit (Amersham Pharmacia Biotech). Before PAGE separation, the focused strips were incubated in 4M urea, 2M thiourea, 30% (v/v) glycerol, 2% (w/v) SDS, 5mM tributyl phosphine 2.5%(w/v) acrylamide, 50mM Tris-HCl pH 8.8, as described [Herbert *et al.* (1998) *Electrophor.* 19:845-51]. SDS-PAGE was performed on linear 9-16% acrylamide gradients. Gels were stained with colloidal Coomassie (Novex, San Diego) [Doherty *et al.* (1998) *Electrophor.* 19:355-63]. Stained gels were scanned with a Personal Densitometer SI (Molecular Dynamics) at 8 bits and 50µm per pixel. Map images were annotated with the software Image Master 2D Elite, version 3.10 (Amersham Pharmacia Biotech). Protein spots were excised from the gel, using an Ettan Spot picker (Amersham Pharmacia Biotech), and dried in a vacuum centrifuge. In-gel digestion of samples for mass spectrometry and extraction of peptides were performed as described by Wilm *et al.* [Nature (1996) 379:466-9]. Samples were desalted with a ZIP TIP (Millipore), eluted with a saturated solution of alpha-cyano-4-hydroxycinnamic acid in 50% acetonitrile, 0.1% TFA and directly loaded onto a SCOUT 381 multiprobe plate (Bruker). Spectra were acquired on a Bruker Biflex II MALDI-TOF. Spectra were calibrated using a combination of known standard peptides, located in spots adjacent to the samples. Resulting values for monoisotopic peaks were used for database searches using the computer program Mascot (www.matrixscience.com). All searches were performed using an error of 200-500ppm as constraint. A representative gel is shown in Figure 190.

#### Example 1

- 35 The following *C.pneumoniae* protein (PID 4376552) was expressed <SEQ ID 1; cp6552>:

1 MKKKLSLLVG LIFVLSSCHK EDAQNKIRIV ASPTPHAELL ESLQEEAKDL

51 GIKLKILPVD DYRIENRLLL DKQVDANYFQ HQAFLLDDECE RYDCKGELVV  
 101 IAKVHLEPQA IYSKHKSSLE RLKSQKKLTI AIPVDRTNAQ RALHLLLEECG  
 151 LIVCKGPANL NMTAKDVCGK ENRSINILEV SAPLLVGSPL DVDAAVIPGN  
 201 FAIAANLSPK KDSLCLLEDLS VSKYTNLVVI RSEDVGSPEM IKLQKLFQSP  
 5 251 SVQHFFDTRY HGNILMTMQD NG\*

A predicted signal peptide is highlighted.

The cp6552 nucleotide sequence <SEQ ID 2> is:

1 ATGAAAAAAA AATTATCATT ACTTGTAGGT TTAATTTTTG TTTTGAGTTC  
 51 TTGCCATAAG GAAGATGCTC AGAATAAAAT ACGTATTGTA GCCAGTCCGA  
 10 101 CACCTCATGC GGAATTATTG GAGAGTTTAC AGGAAGAGGC TAAAGATCTT  
 151 GGAATCAAGC TGAAATACT TCCAGTAGAT GATTATCGTA TTCCTAATCG  
 201 TTTGCTTTTG GATAAACAAG TAGATGCAA TTACTTTCAA CATCAAGCTT  
 251 TTCTTGATGA CGAATGCGAG CGTTATGATT GTAAGGGTGA ATTAGTTGTT  
 301 ATCGCTAAAG TTCATTTGGA ACCTCAAGCA ATTTATTCTA AGAAACATTC  
 15 351 TTCTTTAGAG CGCTTAAAAA GCCAGAAGAA ACTGACTATA GCGATTCCCTG  
 401 TGGATCGTAC GAATGCTCAG CGTGCTCTAC ACTTGTTAGA AGAGTGCCGA  
 451 CTCATTGTTT GCAAAGGGCC TGCTAATTTA AATATGACAG CTAAAGATGT  
 501 CTGTGGGAAA GAAAATAGAA GTATCAACAT ATTAGAGGTG TCAGCTCCTC  
 551 TTCTTGTCGG ATCTCTTCCT GACGTTGATG CTGCTGTCAT TCCTGGAAAT  
 20 601 TTTGCTATAG CAGCAAACCT TTCTCCAAAG AAAGATAGTC TTTGTTTAGA  
 651 GGATCTTTTC GTATCTAAGT ATACAAACCT TGTGTTCATT CGTTCTGAAG  
 701 AGTAGGTTT TCCTAAAATG ATAAAATTAC AGAAGCTGTT TCAATCTCCT  
 751 TCTGTACAAC ATTTTTTGA TACAAAATAT CATGGGAATA TTTTGACAAT  
 801 GACTCAAGAC AATGGTTAG

25 The PSORT algorithm predicts an inner membrane location (0.127).

The protein was expressed in *E.coli* and purified as a his-tag product, as shown in Figure 1A, and also as a GST-fusion. The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 1B) and for FACS analysis (Figure 1C).

The cp6552 protein was also identified in the 2D-PAGE experiment (Cpn0278).

30 These experiments show that cp6552 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

## Example 2

The following *C.pneumoniae* protein (PID 4376736) was expressed <SEQ ID 3; cp6736>:

1 MKTSIRKFLI SFTLAPCFAS TAFTVEVIMP SENFDGSSGK IFPYTTLSDP  
 35 51 RGTLCIFSGD LYIANLDNAI SRTSSSCFSN RAGALQILGK GGVFSFLNIR  
 101 SSADGAAISS VITQNPCLP LSFSGFSQMI FDNCESLTSD TSASNVIPHA  
 151 SAIYATTPML FTNNDLILFQ YNRSAGFGAA IRGTSITIEN TKKSLLFNNG  
 201 GSIISNGGALT GSAAINLINN SAPVIFSTNA TGIYGGATYL TGGSMILTSGN  
 251 LSGVLVFNNS SRSGGAIYAN GNVTFPSNNSD LTFQNTASP QNSLPAPTTP  
 40 301 PTPPAVTPLL GYGGAIFCTP PATPPPTGVS LTISGENSVT FLENIASEQG  
 351 GALYGGKISI DSNKSTIFLG NTAGKGAIA IPESGELSLS ANQGDILFNK  
 401 NLSITSGTPT RNSIHFGKDA KFATLGATQG YTLVYFDPIT SDDLSAASAA  
 451 ATVVVNPKAS ADGAYSGETIV FSGETLTATE AATPANATST LNQKLELEGG  
 501 TLALRNGATL NVHNFTQDEK SVVIMDAGTT LATTNGANNT DGATTLNKLIV  
 45 551 INLDSLDGTR AAVNVQSTN GALTISGTLG LVKNSQDCCD NHGMFNKDLQ  
 601 QVPILKELKAT SNTVTMTDFS LGTNGYQOSP YGYQGTWEFT IDTTHTVTG  
 651 NWKTGGLPH PERLAFLIPN SLWANVIDLR AVSQASAADG EDVPGKQLSI  
 701 TGITNFFHAN HTGDARSYRH MGGGYLINTY TRITPDAALS LGFGQLFTKS  
 751 KDYLVGHHGS NVYFATVYSN ITKSLFGSSR FFSGTSRVT YSRSNEKVKT  
 50 801 SYTKLPKGRG SWSNNCWLGE LEGNLPITLS SRILNLKQII PFVKAERVAYA  
 851 THGGIQENTP EGRIFGHGHL LNVAVFVGVF FGKNSHNRPD FYTIIVAYAP  
 901 DVYRHNPDCE TTPPINGATW TSIGNNLTRS TLLVQASSHT SVNDVLEIFG  
 951 HCGCDIRRTS RQYTLDIGSK LRF\*

A predicted signal peptide is highlighted.



The cp6736 nucleotide sequence <SEQ ID 4> is:

```

1  ATGAAACGT CTATTCGTAA GTTCTTAATT TCTACCACAC TGGCGCCATG
51  TTTTGGCTTCA ACAGCGTTTA CTGTAGAAGT TATCATGCCT TCCGAGAACT
101  TTGATGGATC GAGTGGGAAG ATTTTTCCTT ACACAACACT TTCTGATCCT
5   151  AGAGGGACAC TCTGTATTTT TTCAGGGGAT CTCTACATTG CGAATCTTGA
201  TAATGCCATA TCCAGAACCT CTTCAGTTG CTTTAGCAAT AGGGCGGGAG
251  CACTACAAAT CTTAGGAAAA GGTGGGGTTT TCTCCTTCTT AAATATCCGT
301  TCTTCAGCTG ACGGAGCCGC GATTAGTAGT GTAATCACCC AAAATCCTGA
10  351  ACTATGTCCC TTGAGTTTTC CAGGATTTAG TCAGATGATC TTCGATAACT
401  GTGAATCTTT GACTTCAGAT ACCTCAGCGA GTAATGTCAT ACCTCAGCGA
451  TCGGCGATTT ACGCTACAAC GCCCATGCTC TTTACAAACA ATGACTCCAT
501  ACTATTCCAA TACAACCGTT CTGCAGGATT TGGAGCTGCC ATTCGAGGCA
551  CAAGCATCAC AATAGAAAAA ACGAAAAAGA GCCTTCTCTT TAATGGTAAT
601  GGATCCATCT CTAATGGAGG GGCCCTCAGG GGATCTGCAG CGATCAACCT
15  651  CATCAACAAT AGCGCTCCTG TGATTTTCTC AACGAATGCT ACAGGGATCT
701  ATGGTGGGCG TATTACCTT ACCGGAGGAT CTATGCTCAC CTCTGGGAAC
751  CTCTCAGGAG TCTTGTTCTG TAATAATAGC TCGCGCTCAG GAGGCGCTAT
801  CTATGCTAAC GGAAATGTCA CATTTTCTAA TAACAGCGAC CTGACTTFFC
851  AAAACAATAC AGCATCTCCA CAAAACCTCT TACCTGCACC TACACCTCCA
20  901  CCTACACCAC CAGCAGTCAC TCCTTTGTTA GGATATGGAG GCGCCATCTT
951  CTGTACTCCT CCAGCTACCC CCCCACCAAC AGGTGTTAGC CTGACTATAT
1001  CTGGAGAAAA CAGCGTTACA TTCCTAGAAA ACATTGCCTC CGAACAAAGGA
1051  GGAGCCCTCT ATGGCAAAAA GATCTCTATA GATTCTAATA AATCTACAAT
1101  ATTTCTTGGA AATACAGCTG GAAAAGGAGG CGCTATTGCT ATTCCCGAAT
25  1151  CTGGGGAGCT CTCTCTATCC GCAAATCAAG GTGATATCCT CTTTAAACAAG
1201  AACCTCAGCA TCACTAGTGG GACACCTACT CGCAATAGTA TTCACTTCGG
1251  AAAAGATGCC AAGTTTGCCA CTCTAGGAGC TACGCAAGGC TATACCCTAT
1301  ACTTCTATGA TCCGATTACA TCTGATGATT TATCTGCTGC ATCCCGAGCC
30  1351  GCTACTGTGG TCGTCAATCC CAAAGCCAGT GCAGATGGTG CGTATTTCAGG
1401  GACTATTGTC TTTTCAGGAG AAACCCTCAC TGCTACCGAA GCAGCAACCC
1451  CTGCAAATGC TACATCTACA TTAACCAAAA AGCTAGAACT TGAAGGCGGT
1501  ACTCTCGCTT TAAGAAACGG TGCTACCTTA AATGTTTATA ACTTCACGCA
1551  AGATGAAAG TCCGTCGTCA TCATGGATGC AGGGACCACA TTAGCAACTA
35  1601  CAAATGGAGC TAATAATACT GACGGTGCTA TCACCTTAAA CAAGCTTGTA
1651  ATCAATCTGG ATTCTTTGGA TGGCACTAAA GCGGCTGTCT TTAATGTGCA
1701  GAGTACCAAT GGAGCTCTCA CTATATCCGG AACTTTAGGA CTTGTGAAAA
1751  ACTCTCAAGA TTGCTGTGAC AACCACGGGA TGTTTAATAA AGATTACAG
1801  CAACTTCCGA TTTTAGAACT CAAAGCGACT TCAAATACTG TAACCACTAC
40  1851  GGACTTCAGT CTCGGCACAA ACGGCTATCA GCAATCTCCC TATGGGTATC
1901  AAGGAACTTG GGAGTTTACC ATAGACACGA CAACCCATAC GGTACACAGG
1951  AATTTGAAAA AAACCGGTTA TCTTCTCAT CCGGAGCGTC TTGCTCCCTT
2001  CATTCCTAAT AGCCTATGGG CAAACGTCAT AGATTACGA GCTGTAAGTC
2051  AAGCGTCAGC AGCTGATGGC GAAGATGTCC CTGGGAAGCA ACTGAGCATC
45  2101  ACAGGAATTA CAAATTTCTT CCATGCGAAT CATACCGGTG ATGCACGCAG
2151  CTACCGCCAT ATGGGTGGAG GCTACCTCAT CAATACCTAC ACACGCATCA
2201  CTCCAGATGC TCGTTAAGT CTAGGTTTTC GACAGCTGTT TACAAAATCT
2251  AAGGATTACC TCGTAGGTCA CGGTCACTCT AACGTTTATT TCGCTACAGT
2301  ATACTCTAAC ATCACCAAGT CTCGTGTTGG ATCATCGAGA TTCTTCTCAG
50  2351  GAGGCACCTC TCGAGTTACC TATAGCCGTA GCAATGAGAA AGTAAAGACT
2401  TCATATACAA AATTGCCTAA AGGGCGCTGC TCTTGGAGTA ACAAATGCTG
2451  GTTAGGAGAA CTCGAAGGGA ACCTTCCCAT CACTCTCTCT TCTCGCATCT
2501  TAAACCTCAA GCAGATCATC CCCTTTGTAA AAGCTGAAGT TGCTTACGCG
2551  ACTCATGGGG GCATCCAAGA AAATACCCCC GAGGGGAGGA TTTTGGACA
60  2601  CGGTCACTTA CTCACCGTTG CAGTTCCCGT AGGCGTCCGC TTTGGTAAAA
55  2651  ATTTCTCATA TCGACCAGAT TTTTACACTA TAATCGTAGC CTATGCTCCT
2701  GATGTCTATC GTCAACAATC TGATTGCGAT ACGACATTAC CTATTAATGG
2751  AGCTACGTGG ACCTCTATAG GGAATAATCT AACCAGAAGT ACTTTGCTAG
2801  TACAAGCATC CAGCCATACT TCAGTAAATG ATGTTCTAGA GATCTTCGGG
2851  CACTGTGGAT GTGATATTCG CAGAACCCTC CGTCAATATA CTCTAGATAT
2901  AGGAAGCAAA TTACGATTTT AA

```

The PSORT algorithm predicts an outer membrane location (0.917).

The protein was expressed in *E.coli* and purified as a his-tag product, as shown in Figure 2A, and also as a GST-fusion. Both proteins were used to immunise mice, whose sera were used in a Western blot (Figure 2B) and for FACS analysis (Figure 2C).

The cp6736 protein was also identified in the 2D-PAGE experiment (Cpn0453) and showed good cross-reactivity with human sera, including sera from patients with pneumonitis.

These experiments show that cp6736 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 3

The following *C.pneumoniae* protein (PID 4376751) was expressed <SEQ ID 5; cp6751>:

```

10      1  MRFFCFGMLL PFTFVLANEG LQLPLETYIT LSPBYQAAPO VGFTHNQND
      51  LAIVGNHNDP ILDYKYYSRN GGALTCKNLL ISENIGNVFF EKNVCPNSGG
     101  AIYAAQNCTI SKNQNYAFTT NLVSDNPTAT AGSLLGGALF AINCSITNNL
     151  GQGTFFVDNLA LNKGGALYTE TNLSTKDNKG PIIIKQNRAL NSDSLGGGIY
     201  SGNSLNIBGN SGAIQITSNS SGSGGGIFST QTLTISSNKK LIBEISNSAF
     251  ANNYGSNFPN GGGGLTTTFC TILNNREGVL FNNNQSQSNG GAIHAKSIII
     301  KENGPVYFLN NTATRGGALL NLSAGSGNGS FILSADNGDI IFNNNTASKH
     351  ALNPPYRNAI HSTPNMNLQI GARPGYRVLF YDPIEHELPS SFPILFNFTF
     401  GHTGTIVLFSG EHVHQNFIDE MNFYSYLRNT SELRQGVLA V EDGAGLACYK
     451  FFQRGGTLILL GQGAIVTTAG TIPTPSSTPT TVGSTITLNH IALDLP SILS
     501  FQAQAPKIWI YPTKTGSTYT EDSNPTITIS GTLTLRNSNN EDPYDSL DLS
     551  HSLEKVP LLY IVDVAAQKIN SSQDLSTLN SGEHYGYQGI WSTYVWVETTT
     601  ITNPTSL LGA NTKHKLLYAN WSPLGYRPHP ERRGEFITNA LWQSAYTALA
     651  GLHSLSSWDE EKGAASLQO IGLLVHQKDK NGFRGFRSHM TGYSATTEAT
     701  SSQSPNFSLG FAQFFSKAKE HESQNSTSSH HYFSGMCIEI TLFKEWIRLS
     751  VSLAYMFTSE HTHTMYQGLL EGN SQGSFHN HTLAGALSCV FLPOPHGESL
     801  QIYFPITALA IRGNLAAFQO SGDHAREFSL HRPLTDVSLP VGIRASWKNH
     851  HRVPLVWLTE ISYRSTLYRQ DPELHSKLLI SQGTWTTQAT PVTYNALGIK
     901  VKNVTMQVFPK VTLSLDYSAD ISSSTLSHYL NVASRMRF*

```

A predicted signal peptide is highlighted.

30 The cp6751 nucleotide sequence <SEQ ID 6> is:

```

      1  ATGCCTTTT TTTGCTTCGG AATGTTGCTT CCTTTTACTT TTGTATTGGC
     51  TAATGAAGGT CTCCAAC TTC TTTGGAGAC CTATATTACA TTAAGTCCTG
    101  AATATCAAGC AGCCCTCAA GTAGGGTTTA CTCATAACCA AAATCAAGAT
    151  CTCGCAATTG TCGGGAATCA CAATGATTTC ATCTTGGA CT ATAAGTACTA
    201  TCGGTCGAAT GGAGTGCTC TTACCTGTAA GAATCTCTG ATCTCTGAAA
    251  ATATAGGGAA TGTCTCTCTT GAGAAGAATG TCTGTCCCAA TTCTGGCGGG
    301  GCAATTTATG CTGCTCAAAA TTGCACGATC TCCAAGAATC AGAATATATG
    351  ATTTACTACA AACTTGGTCT CTGACAATCC TACAGCCACT GCGGGATCAC
    401  TATTGGGTGG AGCTCTCTTT GCCATAAAT GCTCTATTAC TAATAACCTA
    451  GGACAGGGAA CTTTCGTTGA CAATCTCGCT TTAATAAGG GGGGTGCCCT
    501  CTATACTGAG ACGAACTTAT CTATTAAAGA CAATAAAGGC CCGATCATAA
    551  TCAAGCAGAA TCGGGCACTA AATTCGGACA GTTTAGGAGG AGGGATTTAT
    601  AGTGGGAAT CTCTAAATAT AGAGGGAAT TCTGGAGCTA TACAGATCAC
    651  AAGCAACTCT TCAGGATCTG GGGGAGGCAT ATTTTCTACC CAAACACTCA
    701  CGATCTCCTC GAATAAAAAA CTCATAGAAA TCAGTGAAAA TTCCGCGTTC
    751  GCAATAACT ATGGATCGAA CTTCAATCCA GGAGGAGGAG GTCTTACTAC
    801  CACCTTTTGC ACGATATTGA ACAACCGAGA AGGGGTACTC TTTAACAATA
    851  ACCAAAGCCA GAGCAACGGT GGAGCCATTC ATCGGAAATC TATCATATATC
    901  AAAGAAAATG GTCTGTATA CTTTPTAAAT AACACTGCAA CTCGGGGAGG
    951  GGCTCTCCTC AACTTATCAG CAGGTCTCTG AAACGGAAGC TTCATCTTAT
   1001  CTGCAGATAA TGGAGATATT ATCTTTAACA ATAATACGGC CTCCAAGCAT
   1051  GCCCTCAATC CTCCATACAG AAACGCCATT CACTCGACTC CTAATATGAA
   1101  TCTGCAATAA GGAGCCGTC CCGGTATCG AGTGCTGTTT TATGATCCCA
   1151  TAGAACATGA GCTCCCTTCC TCCTTCCCCA TACTCTTTAA TTTCGAAACC
   1201  GGTACATACG GTACAGTTTT ATTTTCAGGG GAACATGTAC ACCAGAACTT

```

1251 TACCGATGAA ATGAATTTCT TTTCTATT T AAGGAACACT TCGGAACACT  
 1301 GTCAAGGAGT CCTTGCTGTT GAAGATGGTG CGGGGCTGGC CTGCTATAAG  
 1351 TTCTTCCAAC GAGGAGGCAC TCTACTTCTA GGTCAAGGTG CGGTGATCAC  
 1401 GACAGCAGGA ACGATTCCCA CACCATCCTC AACACCAACG ACAGTAGGAA  
 5 1451 GTACTATAAC TTTAAATCAC ATTGCCATTG ACCTTCCTTC TATTCTTTCT  
 1501 TTTCAGCTC AGGCTCCAAA AATTGGGATT TACCCACAA AAACAGGATC  
 1551 TACCTATACT GAAGATTCCA ACCCGACAAT CACAATCTCA GGAACCTCTCA  
 1601 CCTTACGCAA CAGCAACAAC GAAGATCCCT ACGATAGTCT GGATCTCTCG  
 1651 CACTCTCTTG AGAAAGTTCC CCTTCTTTAT ATTGTCGATG TCGCTGCACA  
 10 1701 AAAAATTAAC TCTTCGCAAC TGGATCTATC CACATTAAAT TCTGGCGAAC  
 1751 ACTATGGGTA TCAAGGCATC TGGTCGACCT ATTGGGTAGA AACTACAACA  
 1801 ATCACGAACC CTACATCTCT ACTAGGCGCG AATACAAAAC ACAAGCTGCT  
 1851 CTATGCAAAC TGGTCTCCTC TAGGCTACCG TCCTCATCCC GAACGTCGAG  
 1901 GAGAATTCAT TACGAATGCC TTGTGGCAAT CGGCATATAC GGCTCTTGCA  
 15 1951 GGAATCCACT CCCCTCTCCTC CTGGGATGAA GAGAAGGGTC ATGCAGCTTC  
 2001 CCTACAAGGC ATTTGGTCTTC TGGTTCATCA AAAAGACAAA AACGGTTTTA  
 2051 AGGGATTTTC TAGTCATATG ACAGGTATTA GTGCTACCAC CGAAGCAACC  
 2101 TCTTCTCAAA GTCCGAATTT CTCTTTAGGA TTTGCTCAGT TCTTCTCCAA  
 2151 AGCTAAAGAA CATGAATCTC AAAATAGCAC GTCCTCTCAC CACTATTTCT  
 20 2201 CTGGAATGTG CATAGAAAAT ACTCTCTTCA AAGAGTGGAT ACGTCTATCT  
 2251 GTGTCTCTTG CTTATATGTT TACCTCGGAA CATACCCATA CAATGTATCA  
 2301 GGGTCTCCTG GAAGGGAAC CTCAGGGATC TTTCCACAAC CATACCTTAG  
 2351 CAGGGGCTCT CTCCTGTGTT TTCCTACCTC AACCTCACGG CGAGTCCCTG  
 2401 CAGATCTATC CCTTTATTAC TGCCTTAGCC ATCCGAGGAA ATCTTGCTGC  
 25 2451 GTTTCAGAA TCTGGAGACC ATGCTCGGGA ATTTTCCCTA CACCGCCCCC  
 2501 TAACGGACGT CTCCCTCCCT GTAGGAATCC GCGCTTCTTG GAAGAACCAC  
 2551 CACCGAGTTC CCCTAGTCTG GCTCACAGAA ATTTCTATC GCTCTACTCT  
 2601 CTATAGGCAA GATCCTGAAC TCCACTCGAA ATTACTGATT AGCCAAGGTA  
 2651 CGTGGACGAC GCAGGCCACT CCTGTGACCT ACAATGCTTT AGGGATCAAA  
 30 2701 GTGAAAAATA CCATGCAGGT GTTTCCTAAA GTCACCTCTC CTTAGATTA  
 2751 CTCTGCGGAT ATTTCTTCTC CCACGCTGAG TCACTACTTA AACGTGGCGA  
 2801 GTAGAATGAG ATTTTAA

The PSORT algorithm predicts an outer membrane location (0.923).

The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 3A, and also in his-tagged form. The GST-fusion recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 3B) and for FACS analysis (Figure 3C).

This protein also showed good cross-reactivity with human sera, including sera from patients with pneumonitis.

These experiments show that cp6751 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

#### Example 4

The following *C.pneumoniae* protein (PID 4376752) was expressed <SEQ ID 7; cp6752>:

1 MFGMTPAVYS LQTDLSLEKFA LERDEEFRTS FPLLDLSLSTL TGFSPITTFV  
 51 GNRHNSQDI VLSNYKSIDN ILLWTSAGG AVSCNNFLLS NVEDHAFFSK  
 45 101 NLAIGTGGAI ACQGACTION NRGPLIFFSN RGLNNASTGG ETRGGAIACN  
 151 GDFTISQNGQ TFYFVNNSVN NWGGALSTNG HCRIQSNRAP LLFFNNTAPS  
 201 GGGALRSNT TISDNTRPIY FKNNCGNNGG AIQTSVTVAI KNNSGSVIFN  
 251 NNTALSGSIN SGNGSGGAIY TTNLSIDNP GTILFNNNYC IRDGGAICTQ  
 301 FLTIKNSGHV YFTNNQGNWG GALMLLDST CLLFAEQGNI AFQNNNEVLT  
 50 351 TFGRYNAIHC TPNSNLQLGA NKGYTTAFD PIEHQHPTTN PLIFNPANAH  
 401 QGTILFSSAY IPEADYENN FISSKNTSE LRNGVLSIED RAGWQFYKFT  
 451 QRGGILKLGH AASIATPANS ETPSTSVGSQ VIINNLAJNL PSILAKGKAP  
 501 TLWIRPLQSS APFTEDNNPT ITLSGPIILL NEENRDPYDS IDLSEPLQNI  
 551 HLLSLSDVTA RHINTDNFHP ESLNATEHYG YQGIWSPYVW ETITTTNNAS  
 55 601 IETANTLYRA LYANWTPLGY KVNPEYQDL ATTPLWQSFH TMFSLLSYN  
 651 RTGDSDIERP FLEIQGIADG LFVHQNSIPG APGFRIQSTG YSLQASSETS

701 LHQKISLGFA QFFTRTKKIG SSNNVSAHNT VSSLYVELPW FQEFATSTV  
 751 LAYGYGDHHL HSLHPSHQEQ AEGTCYSHLT AAAIGCSFPW QQKSYLHLSLSP  
 801 FVQAIARSH QTAFEEIGDN PRKFVSQKPF YNL/TLPLGIQ GKWQSKFHPV  
 851 TEWTLLELSYQ FVLYQQNPQI GVTLLASGGS WDILGHNYVR NALGYKVHVNQ  
 5 901 TALFRSLDLF LDYQGSVSSS TSTHHLQAGS TLKP\*

The cp6752 nucleotide sequence <SEQ ID 8> is:

1 ATGTTCCGGGA TGACTCCTGC AGTGTATAGT TTACAAACGG ACTCCCTTGA  
 51 AAAGTTTGCT TTAGAGAGGG ATGAAGAGTT TCGTACGAGC TTTCCTCTCT  
 101 TAGACTCTCT CTCCACTCTT ACAGGATTTT CTCCAATAAC TACGTTTGTT  
 10 151 GGAAATAGAC ATAATTCCTC TCAAGACATT GTACTTTCTA ACTACAAGTC  
 201 TATGTGATAAC ATCCTTCTTC TTTGGACATC GGCTGGGGGA GCTGTGTCTT  
 251 GTAATAATTT CTTATTATCA AATGTTGAAG ACCATGCCTT CTTCAGTAAA  
 301 AATCTCGCGA TTGGGACTGG AGGCGCGATT GCTTGCCAGG GAGCTGCAC  
 351 AATCAGCAG AATAGAGGAC CCCTTATTTT TTTTCAAGT CGAGGTCTTA  
 15 401 ACAATGCGAG TACAGGAGGA GAAACTCGTG GGGGTGCGAT TGCCTGTAAT  
 451 GGAGACTTCA CGATTTCCTCA AAATCAAGGG ACTTCTTACT TTGTCAACAA  
 501 TTCCGTCAAC AACTGGGGAG GAGCCCTCTC CACCAATGGA CACTGCCGCA  
 551 TCCAAAGCAA CAGGCGACCT CTACTCTTTT TTAACAATAC AGCCCTTAGT  
 601 GGAGGGGGTG CGCTTCGTAG TGAATAACA ACGATCTCTG ATAACACGCG  
 20 651 TCCTATTTAT TTTAAGAACA ACTGTGGGAA CAATGGCGGG GCCATTCAAA  
 701 CAAGCGTTAC TGTGCGGATA AAAATAAAT CCGGTCGGT GATTTTCAAT  
 751 AACAAACACG CGTTATCTGG TTCGATAAAT TCAGGAAATG GTTCAGGAGG  
 801 GCGGATTTAT ACAACAAACC TATCCATAGA CGATAACCTT GGAATATTTC  
 851 TTTTCAATAA TAACTACTGC ATTGCGGATG GCGGAGCTAT CTGTACACAA  
 25 901 TTTTTCACAA TCAAAAATAG TGGCCACGTA TATTTACCA ACAATCAAGG  
 951 AAACCTGGGA GGTGCTCTTA TGCTCTTACA GGACAGCACC TGCCTACTCT  
 1001 TCGCGGAACA AGGAAATATC GCATTTCAA ATAATGAGGT TTTCCTCACC  
 1051 ACATTTGGTA GATAACACGC CATACATTGT ACACCAAATA GCAACTTACA  
 1101 ACTTGGAGCT AATAAGGGGT ATACGACTGC TTTTTTTGAT CCTATAGAAC  
 30 1151 ACCAACATCC AACFACAAAT CCTCTAATCT TTAATCCCAA TGCGAACCAT  
 1201 CAGGGAACGA TCTTATTTTC TTCAGCTTAT ATCCAGAAAG CTCTGACTA  
 1251 CGAAATAAAT TTCATTAGCA GCTCGAAAAA TACCTCTGAA CTTGCAATG  
 1301 GTGTCTCTC TATCGAGGAT CGTGCGGGAT GGCATTTCTA TAAGTTCAT  
 1351 CAAAAGGAG GTATCCTTAA ATTAGGGCAT GCGGCGAGTA TTGCAACAAC  
 35 1401 TGCCCAACTCT GAGACTCCAT CAACFAGTGT AGGCTCCAG GTCATCATTA  
 1451 ATAACCTTGC GATTAACCTC CCCTCGATCT TAGCAAAAGG AAAAGCTCCT  
 1501 ACCTTGTGGA TCCGTCTCTT ACAATCTAGT GCTCTTTTCA CAGAGGACAA  
 1551 TAACCTTACA ATTACTTTAT CAGGTCCTCT GACTCTTTA AATGAGGAAA  
 40 1601 AACGCGATCC CTACGACAGT ATAGATCTCT CTGAGCCTTT ACAAAACATT  
 1651 CATCTTCTTT CTTTATCGGA TGTAACAGCA CGTCATATCA ATACCGATAA  
 1701 CTTTCATCCT GAAAGCTTAA ATGCGACTGA GCATTACGGT TATCAAGGCA  
 1751 TCTGGTCTCC TTATGGGTA GAGACGATAA CAACAACAA TAACGCTTCT  
 1801 ATAGAGACGG CAAACACCTT CTACAGAGCT CTGTATGCCA ATTGGACTCC  
 45 1851 CTTAGGATAT AAGGTCAATC CTGAATACCA AGGAGATCTT GCTACGACTC  
 1901 CCCTATGGCA ATCCTTTTCA ACTATGTTCT CTCTATTAA AAGTTATAAT  
 1951 CGAACTGGTG ATTCTGATAT CGAGAGGCCT TTCTTAGAAA TTCAAGGGAT  
 2001 TGCCGACGGC CTCCTTGTTC ATCAAAATAG CATCCCCGGG GCTCCAGGAT  
 2051 TCCGTATCCA ATCTACAGGG TATTCCTTAC AAGCATCCTC CGAAACTTCT  
 50 2101 TTACATCAGA AAATCTCCTT AGGTTTTGCA CAGTTCTTCA CCCGCACTAA  
 2151 AGAAATCGGA TCAAGCAACA ACGTCTCGGC TCACAATACA GTCTCTTCAC  
 2201 TTTATGTTGA GCTTCCGTGG TTCCAAGAGG CCTTTGCAAC ATCCACAGTG  
 2251 TTAGCGTATG GCTATGGGGA CCATCACCTC CACAGCCTAC ATCCCTCACA  
 2301 TCAAGAACAG GCAGAAGGGA CGTGTATAG CCATACATTA GCAGCAGCTA  
 2351 TCGGCTGTTT TTTCCCTTGG CAACAGAAAT CCTATCTTCA CCTCAGCCCG  
 55 2401 TTCTGTTTCA CAATTGCAAT ACGTTCTCAC CAAACAGCGT TCGAAGAGAT  
 2451 TGGTGACAAT CCCCAGAAAT TTGTCTCTCA AAAGCCTTTC TATAATCTGA  
 2501 CCTTACCTCT AGGAATCCAA GGAAATGGC AGTCAAAAT CCACGTACCT  
 2551 ACAGAAATGA CTCTAGAACT TTCTTACCAA CCGGTACTCT ATCAACAAAA  
 2601 TCCCAAAATC GGTGTACAGC TACTTGCAGG CGGAGGTTC TGGGATATCC  
 60 2651 TAGGCCATAA CTATGTTTCG AATGCTTTAG GGTACAAAGT CCACAAATCA  
 2701 ACTGCGCTCT TCCGTCTCTT CGATCTATTC TTGGATTACC AAGGATCGGT  
 2751 CTCTCTCTCG ACATCTACGC ACCATCTCCA AGCAGGAAGT ACCTTAAAAT  
 2801 TCTAA

The PSORT algorithm predicts a cytoplasmic location (0.138).

The protein was expressed in *E.coli* and purified as a his-tag product, as shown in Figure 4A, and also as a GST-fusion. The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (4B) and the his-tagged protein was used for FACS analysis (4C).

The cp6752 protein was also identified in the 2D-PAGE experiment (Cpn0467).

- 5 These experiments show that cp6752 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 5

The following *C.pneumoniae* protein (PID 4376850) was expressed <SEQ ID 9; cp6850>:

10 1 MKKAVLIAAM FCGVVSLSSC CRIVDCCFED PCAPSSCNPC EVIRKKERS  
51 GGNACGSYVP SCSNFCGSTC CNSQSPQVKG CTSPDGRCKQ \*

A predicted signal peptide is highlighted.

The cp6850 nucleotide sequence <SEQ ID 10> is:

15 1 ATGAAGAAAG CTGTTTAAAT TGCTGCAATG TTTTGTGGAG TAGTTAGCTT  
51 AAGTAGCTGC TGCCGCATTC TAGATTGTTG TTTTGAGGAT CCTTGCGCAC  
101 CCTCTTCTTG CAATCCTTGT GAAGTAATAA GAAAAAAGA AAGATCTTGC  
151 GGCCTAATG CTTGTGGGTC CTACGTTCTT TCTTGTCTTA ATCCATGTGG  
201 TTCAACAGAG TGTAACTCTC AAAGCCACA AGTTAAAGGT TGTACATCAC  
251 CTGATGGCAG ATGCAAACAG TAA

The PSORT algorithm predicts an inner membrane location (0.329).

- 20 The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 5A. The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 5B) and for FACS analysis (Figure 5B). A his-tagged protein was also expressed.

These experiments show that cp6850 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### 25 Example 6

The following *C.pneumoniae* protein (PID 4376900) was expressed <SEQ ID 11; cp6900>:

30 1 MKIKFSWKVN FLICLLAVGL IFFGCSRVKR EVLVGRDATW FPKQFGIYTS  
51 DTNAFLNDLV SEINYKENLN INIVNQDWVH LFENLDDKRT QGAFTSVLPT  
101 LEMLEHYQFS DPILLTGPFV VVAQDSPYQS IEDLGRLIG VYKFDSSVLV  
151 AQNIPDAVIS LYQHVPIALE ALTSNCYDAL LAPVIEVTAL IETAYKGRLE  
201 IISKPLNADG LRLAILKGTN GDLLEGFNAG LVKTRRSKY DAIRQYRLP

The cp6900 nucleotide sequence <SEQ ID 12> is:

35 1 GTGAAGATAA AATTTTCTTG GAAGGTAAAT TTTTAAATAT GTTACTGGC  
51 TGTGGGACTG ATCTTTTTCG GGTGCTCTCG AGTAAAAAGA GAAGTTCTCG  
101 TAGGTCGTGA TGCCACCTGG TTTCAAAAC AATTCGGCAT TTATACATCC  
151 GATACCAACG CATTTTAAAC CGATCTTGTT TCTGAGATTA ACTATAAAGA  
201 GAATCTAAAT ATTAATATTG TAAATCAAGA TTGGGTGCAT CTCCTTGAGA  
251 ATTTAGATGA TAAAAGACC CAAGGAGCAT TTACATCTGT ATTGCCTACT  
301 CTGAGATGC TCGAACACTA TCAATTTTCT GATCCCATTT TACTCACAGG  
40 351 TCCGTGCTCT GTGTCGCTC AAGACTCTCC TTACCAATCT ATAGAGGATC  
401 TTAAGGTCG TCTTATTGGA GTGTATAAGT TTGACTCTTC AGTCTTGTA  
451 GCTCAAATA TCCCTGACGC TGTGATTAGC CTCTACCAAC ATGTTCCAAT  
501 AGCATTGGAA GCCTTAACAT CGAATTGTTA CGACGCTCTT CTAGCTCCTG  
551 TAATTGAAGT GACCGCGCTA ATAGAAACAG CATATAAAGG AAGACTGAAA  
45 601 ATTATTTCAA AACCTTAAA CGCAGATGGT TTGCGGCTTG CAATACTGAA

-47-

651 AGGGACAAAC GGAGATTTGC TTGAAGGGTT TAACGCAGGA CTTGTGAAAA  
 701 CACGACGCTC AGGAAAATAC GATGCTATAA AACAGCGGTA TCGTCTTCCC  
 751 TAA

The PSORT algorithm predicts an inner membrane location (0.452).

- 5 The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 6A. The recombinant protein was used to immunise mice, whose sera were used for FACS analysis (Figure 6B). A his-tagged protein was also expressed.

The cp6900 protein was also identified in the 2D-PAGE experiment (Cpn0604).

- 10 These experiments show that cp6900 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 7

The following *C.pneumoniae* protein (PID 4377033) was expressed <SEQ ID 13; cp7033>:

1 MVNPIGPGPI DETERTPPAD LSAQGLEASA ANKSAEAQRI AGAEAKPKES  
 51 KTDSEVERWSI LRSVNALMS LADKLGIASS NSSSTSRSA DVDSTTATAP  
 101 TPPPPTFFDY KTQQAAYDT IFTSTSLADI QAALVSLQDA VTNIKDTAAT  
 151 DEETAIAAEW ETKNADAVKV GAQITELAKY ASDNQAILDS LGKLTFSFDLL  
 201 QAALLQSVAN NNKAAELLKE MQDNPVVPKG TPAIAQSLVD QTDATATQIE  
 251 KDGNAIRDAY FAGQNASGAV ENAKSNNSIS NIDSAKAAIA TAKTQIAEAQ  
 301 KKFDPSPILQ RAEQMVIQAE KDLKNIKPAD GSDVPNPGTT VGGSKQQGSS  
 20 351 IGSIRVSMIL DDAENETASI LMSGFRQMIH MFNTENPDSQ AAQQLAAQA  
 401 RAAKAAGDDS AAAALADAQK ALEAALGKAG QQQGILNALG QIASAAVSA  
 451 GVPPAAASSI GSSVKQLYKT SKSTGSDYKT QISAGYDAYK SINDAYGRAR  
 501 NDATRDVINN VSTPALTRSV PRARTEARGP EKTDQALARV ISGNSRTLGD  
 551 VYSQVSALQS VMQIIQSNPQ ANNEEIROKL TSAVTKPPQF GYPYVQLSND  
 25 601 STQKFIKLE SLFAEGSRTA AEIKALSFET NSLFIQQVLV NIGSLYSGYL  
 651 Q\*

The cp7033 nucleotide sequence <SEQ ID 14> is:

1 ATGGTTAATC CTATTGGTCC AGGTCCTATA GACGAAACAG AACGCACACC  
 51 TCCCGCAGAT CTITCTGCTC AAGGATTGGA GCGAGTGCA GCAAATAAGA  
 30 101 GTGCGGAAGC TCAAAGAATA GCAGGTGCGG AAGCTAAGCC TAAAGAATCT  
 151 AAGACCGATT CTGTAGAGCG ATGGAGCATC TTGCGTTCTG CAGTGAATGC  
 201 TCTCATGAGT CTGCGAGATA AGCTGGGTAT TGCTTCTAGT AACAGCTCGT  
 251 CTCTACTAG CAGATCTGCA GACGTGGACT CAACGACAGC GACCGCACCT  
 301 ACGCTCCTC CACCCACGTT TGATGATTAT AAGACTCAAG CGCAAACAGC  
 35 351 TTACGATACT ATCTTTACCT CAACATCACT AGCTGACATA CAGGCTGCTT  
 401 TGGTGAGCCT CCAGGATGCT GTCATAATA TAAAGGATAC AGCGCTACT  
 451 GATGAGGAAA CCGCAATCGC TCGGAGGTGG GAAACTAAGA ATGCCGATGC  
 501 AGTTAAAGTT GCGCGCAAA TTACAGAATT AGCGAAATAT GCTTCGGATA  
 551 ACCAAGCGAT TCTTGACTCT TTAGGTAAAC TGACTTCCTT CGACCTCTTA  
 40 601 CAGGCTGCTC TTCTCCAATC TGTAGCAAAC AATAACAAAG CAGCTGAGCT  
 651 TCTTAAAGAG ATGCAAGATA ACCCAGTAGT CCCAGGGAAA ACGCTGCAA  
 701 TTGCTCAATC TTTAGTTGAT CAGACAGATG CTACAGCGAC ACAGATAGAG  
 751 AAAGATGGAA ATGCGATTAG GGATGCATAT TTTGCAGGAC AGAACGCTAG  
 801 TGGAGCTGTA GAAATGCTA AATCTAATAA CAGTATAAGC AACATAGATT  
 45 851 CAGCTAAAGC AGCAATCGCT ACTGCTAAGA CACAAATAGC TGAAGCTCAG  
 901 AAAAAGTTCC CCGACTCTCC AATCTTTCAA GAAGCGGAAC AAATGGTAAT  
 951 AAGAGCTGAG AAAGATCTTA AAAATATCAA ACCTGCAGAT GTTCTGATG  
 1001 TTCCAAATCC AGGAATACA GTTGGAGGCT CCAAGCAACA AGGAAGTAGT  
 1051 ATTGGTAGTA TTCGTGTTTC CATGCTGTTA GATGATGCTG AAAATGAGAC  
 50 1101 CGCTTCCATT TTGATGCTCG GTTTCGTCA GATGATTCAC ATGTTCAATA  
 1151 CGGAAAATCC TGATTCTCAA GCTGCCAAC AGGAGCTCGC AGCACAAGCT  
 1201 AAGAGCAGCA AAGCCGCTGG AGATGACAGT GCTGCTGCAG CGCTGGCAGA  
 1251 TGCTCAGAAA GCTTTAGAAG CGGCTCTAGG TAAAGCTGGG CAACAACAGG  
 1301 GCATACTCAA TGCTTTAGGA CAGATCGCTT CTGCTGCTGT TGTGAGCGCA  
 55 1351 GGAGTTCCTC CCGCTGCAGC AAGTTCCTATA GGGTCATCTG TAAACAGCT  
 1401 TTACAAGACC TCAAAATCTA CAGGTTCTGA TTATAAAACA CAGATATCAG

1451 CAGGTTATGA TGCTTACAAA TCCATCAATG ATGCCTATGG TAGGGCACGA  
 1501 AATGATGCGA CTCGTGATGT GATAAACAAAT GTAAGTACCC CCGCTCTCAC  
 1551 ACGATCCGTT CCTAGAGCAC GAACAGAAGC TCGAGGACCA GAAAAAACAG  
 1601 ATCAAGCCCT CGCTAGGGTG ATTTCTGGCA ATAGCAGAAC TCTTGGAGAT  
 5 1651 GTCTATAGTC AAGTPTCGGC ACTACAATCT GTAATGCAGA TCATCCAGTC  
 1701 GAATCCTCAA GCGAATAATG AGGAGATCAG ACAAAGCTT ACATCGGCAG  
 1751 TGACAAAGCC TCCACAGTTT GGCTATCCTT ATGTGCAACT TTCTAATGAC  
 1801 TCTACACAGA AGTTCATAGC TAAATAGAA AGTTTGTGTTG CTGAAGGATC  
 1851 TAGGACAGCA GCTGAAATAA AAGCACTTTC CTTTGAAACG AACTCCTTGT  
 10 1901 TTATTCAGCA GGTGCTGGTC AATATCGGCT CTCTATATTC TGGTTATCTC  
 1951 CAATAA

The PSORT algorithm predicts a cytoplasmic location (0.272).

The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 7A. A  
 his-tagged protein was also expressed. The recombinant proteins were used to immunise mice, whose  
 15 sera were used for FACS (Figure 7B) and Western blot (7C) analyses.

The cp7033 protein was also identified in the 2D-PAGE experiment (Cpn0728) and showed good cross-reactivity with human sera, including sera from patients with pneumonitis.

These experiments show that cp7033 a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

## 20 Example 8

The following *C.pneumoniae* protein (PID 6172321) was expressed <SEQ ID 15; cp0017>:

1 MGIKGTGIIV WVDDATAKTK NATLITWTKTG YKPNPERQGP LVPNSLWGSF  
 51 VDVRSIQSLM DRSTSSLSSS TNLWVSGIAD FLHEDQKGNQ RSYRHSSAGY  
 101 ALGGGFFTAS ENFFNFAFCQ LFGYDKDHLV AKNHHTVYAG AMSYRHLGES  
 25 151 KTLAKILSGN SDSLPFVFNA RFAYGHTDNN MTTKYTGYSY VKGSWGNDAF  
 201 GIECGGAIPV VASGRRSWVD THTPFLNLEM IYAHQNDPKE NGTEGRSFQS  
 251 EDLFLNLAIPV GIKFEKFSK STYDLNLAIV PDVIRNDPGC TTTLMVSGDS  
 301 WSTCGTSLSR QALLVRAGNH HAFASNFVVF SQFEVELRGS SRSYAIDLGG  
 351 RFGF\*

30 The cp0017 nucleotide sequence <SEQ ID 16> is:

1 ATGGGTATCA AGGGAACCTG AATAATTGTT TGGGTCGACG ATGCAACTGC  
 51 AAAAAACAAAA AATGCTACCT TAACCTGGAC TAAAACAGGA TACAAGCCGA  
 101 ATCCAGAACG TCAGGGACCT TTGGTTCCTA ATAGCCTGTG GGGTTCCTTT  
 35 151 GTCGATGTCC GCTCCATTCA GAGCCTCATG GACCGGAGCA CAAGTTCGTT  
 201 ATCTTCGTCA ACAAATTGTT GGGTATCAGG AATCGCGGAC TTTTTCGATG  
 251 AAGATCAGAA AGGAAACCAA CGTAGTTATC GTCATTCTAG CGCGGGTTAT  
 301 GCATTAGGAG GAGGATTCCT CACGGCTTCT GAAAATTCTT TTAATTTTGC  
 351 TTTTGTGTCAG CTTTGTGGCT ACGACAAGGA CCATCTTGTC GCTAAGAACCC  
 401 ATACCCATGT ATATGCAAGG GCAATGAGTT ACCGACACCT CGGAGAGTCT  
 451 AAGACCCTCG CTAAGATTTT GTCAGGAAAT TCTGACTCCC TACCTTTTGT  
 501 CTTCAATGCT CGGTTTGCTT ATGGCCATAC CGACAATAAC ATGACCACAA  
 551 AGTACACTGG CTATCTCTCT GTTAAGGGAA GCTGGGGAAA TGATGCCTTC  
 601 GGTATAGAAT GTGGAGGAGC TATCCCGGTA GTTGCTTCAG GACGTCGGTC  
 651 TTGGGTGGAT ACCCACACGC CATTTCCTAAA CCTAGAGATG ATCTATGCAC  
 45 701 ATCAGAAATG CTTTAAGGAA AACGGCACAG AAGGCCGTTT TTTCCAAAGT  
 751 GAAGACCTCT TCAATCTAGC GGTTCCTGTA GGGATAAAAT TTGAGAAATT  
 801 CTCCGATAAG TCTACGTATG ATCTCTCCAT AGCTTACGTT CCCGATGTGA  
 851 TTCGTAATGA TCCAGGCTGC ACGACAACCT TTATGGTTTC TGGGGATTCT  
 901 TGGTCGACAT GTGGTACAAG CTTGTCTAGA CAAGCTCTTC TTGTACGTGC  
 50 951 TGGAAATCAT CATGCCTTTG CTTCAAACCT TGAAGTTTTC AGTCAGTTTG  
 1001 AAGTCGAGTT GCGAGGTTCT TCTCGTAGCT ATGCTATCGA TCTTGGAGGA  
 1051 AGATTCGGAT TTAA

This sequence is frame-shifted with respect to cp0016.

The PSORT algorithm predicts a cytoplasmic location (0.075).

The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 8A. The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 8B) and for FACS analysis (Figure 8C). A his-tagged protein was also expressed.

- 5 This protein also showed good cross-reactivity with human sera, including sera from patients with pneumonitis.

These experiments show that cp0017 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

#### Example 9

- 10 The following *C.pneumoniae* protein (PID 6172315) was expressed <SEQ ID 17; cp0014>:

```

1  MKSSPFRFVF STFAIFPLSM IATETVLDSS ASFDGNKNGN FSVRESQEDA
51  GTTYLFKGNV TLENIPGTGT AITKSCFNNT KGDLTFTGNG NSLLFQTVDA
101 GTVAGAAVNS SVVDKSTTFI GFSSLSFIAS PGSSITTGKG AVSCSTGSLs
151 LTKMSVCSSA KTFQRIMAVL SPQKLFH*
```

- 15 The cp0014 nucleotide sequence <SEQ ID 18> is:

```

1  ATGAAGTCIT CTTCCTCCAA GTTTGTATTT TCTACATTG CTATTTTCCC
51  TTTGTCTATG ATTGCTACCG AGACAGTTT GGATTCAGT GCGAGTTTCG
101 ATGGGAATAA AAATGGTAAT TTTTCAGTTC GTGAGAGTCA GGAAGATGCT
20  GGAAGTACCT ACCTATTTAA GGGAAATGTC ACTCTAGAAA ATATTCCTGG
201 AACAGGCACA GCAATCACAA AAAGCTGTTT TAACAACACT AAGGGCGATT
251 TGACTTTTAC AGGTAAACGGG AACTCTCTAT TGTTCCAAAC GGTGGATGCA
301 GGGACTGTAG CAGGGGCTGC TGTAAACAGC AGCGTGGTAG ATAAATCTAC
351 CACGTTTATA GGGTTTCTTT CGCTATCTTT TATTGCGTCT CCTGGAAGTT
25  CGATAACTAC CGGCAAAGGA GCCGTTAGCT GCTCTACGGG TAGCTTGAGT
451 TTGACAAAAA TGTCAGTTTG CTCCTCAGCA AAAACTTTTC AACGGATAAT
501 GCGGGTGCTA TCACCGCAAA AACTCTTTCA TTAA
```

This protein is frame-shifted with respect to cp0015.

The PSORT algorithm predicts an inner membrane location (0.047).

- 30 The protein was expressed in *E.coli* and purified as a his-tag product, as shown in Figure 9A. A GST-fusion was also expressed. The recombinant proteins were used to immunise mice, whose sera were used in an immunoassay (Figure 9B) and for FACS analysis (Figure 9C).

This protein also showed good cross-reactivity with human sera, including sera from patients with pneumonitis.

- 35 These experiments suggest that cp0014 is a useful immunogen. These properties are not evident from the sequence alone.

#### Example 10

The following *C.pneumoniae* protein (PID 6172317) was expressed <SEQ ID 19; cp0015>:

```

1  MSALFSENTS SKKGAIQTS DALTTTGNQG EVSFDNTSS DSGAALFTEA
40 51  SVTISNNAKV SFIDNKVTGA SSSTTGDMSS GAICAYKTST DTKVTLTGNQ
101 MLLFSNNTST TAGGAIYVKK LELASGGLTL FSRNSVNGGT APKGGAIAIE
151 DSGELSLSD SGDIVFLGNT VTSTTPGTNR SSIDLGTSK MTALRSAAGR
```



-50-

201 AIIFYDPIIT GSSTTVTDVL KVNETPADSA LQYTGNIIFT GEKLSETEAA  
 251 DSKNLTSLKLL QPVTLSSGGL SLKHGVTLLQT QAFYQQADSR LEMDVGTTLLE  
 301 PADTSTINNL VINISSIDGA KKAKIETKAT SKNLTLSGTI TLLDPTGTFY  
 5 351 ENHSLRNPOQS YDILELKASG TVTSTAVTPD PIMGEKFFHYG YQGTWGPVW  
 401 GTGASTTATF NWTKTGYIPN PERIGSLVPN SLWNAFIDIS SLHYLMETAN  
 451 EGLQGDRAFW CAGLSNFFHK DSTKTRRGFR HLSGGYVIGG NLHTCSDKIL  
 501 SAAFCQLFGR DRDYFVAKNQ GTVYGGTLYY QHNETYISLP CKLRPCSLSY  
 551 VPTEIPVLFS GNLSYTHTDN DLKTKYTTYP TVKGSWGNDS FALBFGGRAP  
 601 ICLDESALFE QYMPFMKLQF VYAHQEGPKE QGTEAREFGS SRLVNLALPI  
 10 651 GIRFDKESDC QDATYNLTG YTVDLVRSNP DCTTTLRISG DSWKTFGTNL  
 701 ARQALVLRAG NHFCFNSNFE AFSQFSFELR GSSRNYNVDL GAKYQF\*

This sequence is frame-shifted with respect to cp0014.

The cp0015 nucleotide sequence <SEQ ID 20> is:

1 ATGTCAGCTC TGTTTTCTGA AAATACCTCC TCAAAGAAAG GCGGAGCCAT  
 15 51 TCAGACTTCC GATGCCCTTA CCATTACTGG AAACCAAGGG GAAGTCTCTT  
 101 TTTCTGACAA TACTTCTTCG GATTCTGGAG CTGCAATTTT TACAGAAGCC  
 151 TCGGTGACTA TTTCTAATAA TGCTAAAGTT TCCTTTATTG ACAATAAGGT  
 201 CACAGGAGCG AGCTCCTCAA CAACGGGGGA TATGTCAGGA GGTGCTATCT  
 251 GTGCTTATAA AACTAGTACA GATACTAAGG TCACCCTCAC TGGAAATCAG  
 20 301 ATGTTACTCT TCAGCAACAA TACATCGACA ACAGCGGGAG GAGCTATCTA  
 351 TGTGAAAAAG CTCGAACTGG CTTCCGGAGG ACTTACCCTA TTCAGTAGAA  
 401 ATAGTGTCAA TGGAGGTACA GCTCCTAAAG GTGGAGCCAT AGCTATCGAA  
 451 AGATGTGGGG AATTGAGTTT ATCCGCCGAT AGTGGTGACA TTGCTTTTTT  
 501 AGGGAATACA GTCACCTCTA CTACTCCTGG GACGAATAGA AGTAGTATCG  
 25 551 ACTTAGGAAC GAGTGCAAAAG ATGACAGCTT TGCCTTCTGC TGCTGGTAGA  
 601 GCCATCTACT TCTATGATCC CATAACTACA GGATCATCCA CAACAGTTAC  
 651 AGATGTCTTA AAAGTTAATG AGACTCCGGC AGATTCTGCA CTACAATATA  
 701 CAGGGAACAT CATCTTCACA GGAGAAAAGT TATCAGAGAC AGAGGCCGCA  
 751 GATTCTAAAA ATCTTACTTC GAAGCTACTA CAGCTGTAA CTCCTTCAGG  
 30 801 AGGTACTCTA TCTTTAAAC ATGGAGTGAC TCTGCAGACT CAGGCATTCA  
 851 CTCACAGGC AGATTCTCGT CTCGAAATGG ACGTAGGAAC TACTCTAGAA  
 901 CCTGCTGATA CTAGCACCAT AAACAATTTG GTCATTAAACA TCAGTTCTAT  
 951 AGACGGTGCA AAGAAGGCAA AAATAGAAAC CAAAGCTACG TCAAAAATC  
 1001 TGACTTTATC TGGAAACCATC ACTTTATTGG ACCCGACGGG CACGTTTAT  
 35 1051 GAAAATCATA GTTTAAGAAA TCCTCAGTCC TACGACATCT TAGAGCTCAA  
 1101 AGCTTCTGGA ACTGTAAACAA GCACCGCAGT GACTCCAGAT CCTATAATGG  
 1151 GTGAGAAAT CCATTACGGC TATCAGGGAA CTTGGGGCCC AATTGTTTGG  
 1201 GGGACAGGGG CTTCTACGAC TGCAACCTTC AACTGGACTA AAACGGCTA  
 1251 TATTCCTAAT CCCGAGCGTA TCGGCTCTTT AGTCCCTAAT AGCTTATGGA  
 40 1301 ATGCATTTAT AGATATTAGC TCTCTCCATT ATCTTATGGA GACTGCAAAAC  
 1351 GAAGGGTTGC AGGGAGACCG TGCTTTTGG TGTGCTGGAT TATCTAACTT  
 1401 CTTCCATAAG GATAGTACAA AAACACGACG CGGGTTTCGC CATTTGAGTG  
 1451 CGGGTTATGT CATAGGAGGA AACCTACATA CTTGTTTCTA TAAGATTCTT  
 45 1501 AGTGTGTCAT TTTGTGAGCT CTTTGGGAAGA GATAGAGACT ACTTTGTAGC  
 1551 TAAGAAATCAA GGTACAGTCT ACGGAGGAAC TCTCTATTAC CAGCACAACG  
 1601 AAACCTATAT CTCTCTTCTT TGCAAACTAC GGCCTTGTTC GTTGTCTTAT  
 1651 GTTCCTACAG AGATTCCTGT TCTCTTTTCA GGAAACCTTA GCTACACCCA  
 1701 TACGGATAAC GATCTGAAAA CCAAGTATAC AACATATCCT ACTGTTAAAG  
 1751 GAAGCTGGGG GAATGATAGT TTCGCTTTAG AATTCCGGTGG AAGAGCTCCG  
 50 1801 ATTTGCTTAG ATGAAAGTGC TCTATTTGAG CAGTACATGC CCTTCATGAA  
 1851 ATTGCAAGTTT GTCTATGCAC ATCAGGAAGG TTTTAAAGAA CAGGGAACAG  
 1901 AAGCTCGTGA ATTTGGAAGT AGCCGTCTTG TGAATCTTGC CTTACCTATC  
 1951 GGGATCCGAT TTGATAAGGA ATCAGACTGC CAAGATGCAA CGTACAATCT  
 2001 AACTCTTGGT TATACTGTGG ATCTTGTTCG TAGTAACCCC GACTGTACGA  
 55 2051 CAACACTGCG AATTAGCGGT GATTCTTGGG AAACCTTCGG TACGAATTTG  
 2101 GCAAGACAAG CTTTAGTCTT TCGTGACGGG AACCAATTTT GCTTTAACTC  
 2151 AAATTTTGA GCCTTTAGCC AATTTCTTTT TGAATTGCGT GGGTCATCTC  
 2201 GCAATTACAA TGTAGACTTA GGAGCAAAAT ACCAATTCTA A

The PSORT algorithm predicts a cytoplasmic location (0.274).

60 The protein was expressed in *E. coli* and purified as a GST-fusion product, as shown in Figure 10A. The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 10B) and for FACS analysis. A his-tagged protein was also expressed.

These experiments show that cp0015 is a useful immunogen. These properties are not evident from the sequence alone.

### Example 11

The following *C.pneumoniae* protein (PID 6172325) was expressed <SEQ ID 21; cp0019>:

```

5      1  LQDSQDYSFV  KLSPGAGGTI  ITQDASQKPL  EVAPSRPHYG  YQGHWNVQVI
      51  PGTGTQPSQA  NLEWVRTGYL  PNPERQGSV  PNSLWGSFVD  QRAIQEIMVN
     101  SSQILQERG  VWGAGIANFL  HRDKINEHGY  RHSGVGYLVG  VGTHAFSDAT
     151  INAAFCQLFS  RDKDYVVSKN  HGTSYSGVVF  LEDTLEFRSP  QGFYTDSSSE
     201  ACCNQVVTID  MQLSYSHRNN  DMKTKYTTY  EAQGSWANDV  FGLEFGATTY
    10  251  YYPNSTFLFD  YYSFPLRLQC  TYAHQEDFKE  TGGEVRHFTS  GDLFNLAVPI
     301  GVKPERFSDC  KRGSYELTLA  YVPDVIRKDP  KSTATLASGA  TWSTHGNNLS
     351  RQLQLRLGN  HCLINPGIEV  FSHGAIELRG  SSRNYNINLG  GKRYRF*

```

This sequence is frame-shifted with respect to cp0018.

The cp0019 nucleotide sequence <SEQ ID 22> is:

```

15      1  TTGCAAGACT  CTCAAGACTA  TAGCTTTGTA  AAGTTATCTC  CAGGAGCGGG
      51  AGGGACTATA  ATTACTCAAG  ATGCTTCTCA  GAAGCCTCTT  GAAGTAGCTC
     101  CTTCTAGACC  ACATTATGGC  TATCAAGGAC  ATTGGAATGT  GCAAGTCATC
     151  CCAGGAACGG  GAACTCAACC  GAGCCAGGCA  AATTTAGAAT  GGGTGCGGAC
     201  AGGATACCTT  CCGAATCCCG  AACGGCAAGG  ATCTTTAGTT  CCCAATAGCC
    20  251  TGTGGGGTTC  TTTTGTGTAT  CAGCGTGCTA  TCCAAGAAAT  CATGGTAAAT
     301  AGTAGCCAAA  TCTTATGTCA  GGAACGGGGA  GTCTGGGGAG  CTGGAATTGC
     351  TAATTTCTTA  CATAGAGATA  AAATTAATGA  GCACGGCTAT  CGCCATAGCG
     401  GTGTCGGTTA  TCTTGTGGGA  GTTGGCACTC  ATGCTTTTTC  TGATGCTACG
     451  ATAAATGCGG  CTTTGTGCCA  GCTCTTCAGT  AGAGATAAAG  ACTACGTAGT
    25  501  ATCCAAAAAT  CATGGAACCT  GCTACTCAGG  GGTGCTATTT  CTTGAGGATA
     551  CCCTAGAGTT  TAGAAGTCCA  CAGGGATTCT  ATACTGATAG  CTCTCAGAA
     601  GCTTGCTGTA  ACCAAGTCGT  CACTATAGAT  ATGCAGTTGT  CTTACAGCCA
     651  TAGAAATAAT  GATATGAAAA  CCAAATACAC  GACATATCCA  GAAGCTCAGG
     701  GATCTTGGGC  AAATGATGTT  TTTGGTCTTG  AGTTTGGAGC  GACTACATAC
    30  751  TACTACCCTA  ACAGTACTTT  TTTATTTGAT  TACTACTCTC  CGTTTCTCAG
     801  GCTGCAGTGC  ACCTATGCTC  ACCAGGAAGA  CTTCAAAGAG  ACAGGAGGTG
     851  AGGTTTCGTC  CTTTACTAGC  GGAGATCTTT  TCAATTTAGC  AGTTCCTATT
     901  GCGGTGAAGT  TTGAGAGATT  TTCAGACTGT  AAAAGGGGAT  CTTATGAACT
     951  TACCCTTGCT  TATGTTCTTG  ATGTGATTCT  CAAAGATCCC  AAGAGCACGG
    35  1001  CAACATTGGC  TAGTGGAGCT  ACGTGGAGCA  CCCACGGAAA  CAATCTCTCC
     1051  AGACAAGGAT  TACAACCTGC  TTTAGGGAAC  CACTGTCTCA  TAAATCCTGG
     1101  AATTGAGGTG  TTCAGTCACG  GAGCTATTGA  ATTGCGGGGA  TCCTCTCGTA
     1151  ATTATAACAT  CAATCTCGGG  GGTAAATACC  GATTTTAA

```

The PSORT algorithm predicts a cytoplasmic location (0.189).

40 The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 11A. This protein was used to immunise mice, whose sera were used in a Western blot (Figure 11B) and an immunoblot assay (Figure 11C). A his-tagged protein was also expressed.

These experiments show that cp0019 is a useful immunogen. These properties are not evident from the sequence alone.

### 45 Example 12

The following *C.pneumoniae* protein (PID 4376466) was expressed <SEQ ID 23; cp6466>:

```

      1  MRKISVGICI  TILLSLSVVL  QGCKESSHSS  TSRGELAINI  RDEPRSLDPR
     51  QVRLLSEISL  VKHYEGLVQ  ENNLSGNIEP  ALAEDYSLSS  DGLTYTFKLR
    101  SAFWSNGDPL  TAEDFIESWK  QVATQEVSGI  YAFALNPIKN  VRKIQEGHLS
    151  IDHFGVHSPN  ESTLVVTLES  PTSHFLKLLA  LPVFPVHKS  QRTLQSKSLP
    201  IASGAFYPKN  IKQKQWIKLS  KNPHYYNQSQ  VETKTITIH  IPDANTAACL

```

251 FNQGLNMQG PPWGERIPQE TLSNLQSKGH LHSFDVAGTS WLTFNINKFP  
 301 LNNMKLREAL ASALDKEALV STIFLGRAKT ADHLLPTNIH SYPEHQKQEM  
 351 AQRQAYAKKL FKEALEELQI TAKDLEHLNL IFPVSSASS LLVQLIREQW  
 401 KESLGFAIPI VGKEFALLQA DLSSGNFSLA TGGWFADFAD PMAFLTIFAY  
 5 451 PSGVPPYAIN HKDFLEILQN IEQEQDHQKR SELVSQASLY LETFHIIEPI  
 501 YHDAFQFAMN KKLNLGVSP TGVVDFRYAK EN\*

A predicted signal peptide is highlighted.

The cp6466 nucleotide sequence <SEQ ID 24> is:

10 1 ATGCGCAAGA TATCAGTGGG AATCTGTATC ACCATTCTCC TTAGCCTCTC  
 51 CGTAGTCTCT CAAGGCTGCA AGGAGTCCAG TCACTCCTCT ACATCTCGGG  
 101 GAGAACTCGC TATTAATATA AGAGATGAAC CCCGTCTCTT AGATCCAAGA  
 151 CAAGTGCAGC TTCTTTCAGA AATCAGCCTT GTCAAACATA TCTATGAGGG  
 201 ATTAGTTCAA GAAATAATC TTTCAGGAAA TATAGAGCCT GCTCTTGCAG  
 15 251 AAGACTACTC TCCTTCCTCG GACGACTCA CTTATACTTT TAAACTGAAA  
 301 TCAGCTTTT GAGTAATGG CGACCCCTTA ACAGCTGAAG ACTTTATAGA  
 351 ATCTTGGAAA CAAGTAGCTA CTCAGGAAGT CTCAGGAATC TATGCTTTTG  
 401 CCTTGAATCC AATFAAAAT GTACGAAAGA TCCAAGAGGG ACACCTCTCC  
 451 ATAGACCAAT TTGGAGTGCA CTCCTCTAAT GAATCTACAC TTGTGTGTAC  
 501 CCTGGAATCC CCAAGCTCGC ATTTCTTAAA ACTTTTAGCT CTTCAGTCT  
 20 551 TTTTCCCCGT TCATAAATCT CAAAGAACCC TGCAATCCAA ATCTCTACCT  
 601 ATAGCAAGCG GAGCTTTCTA TCCTAAAAAT ATCAAAACAA AACAATGGAT  
 651 AAAACTCTCA AAAAACCTC ACTACTATAA TCAAAGTCAG GTGGAACTA  
 701 AAACGATTAC GATTCACTTC ATTCCCGATG CAAACACAGC AGCAAACTA  
 751 TTTAATCAGG GAAAACCTCA TTGGCAAGGA CCTCTTGGG GAGAACGCAT  
 25 801 TCCTCAAGAA ACCCTATCCA ATTTACAGTC TAAGGGGCAC TTACTACTCTT  
 851 TTGATGTGCG AGGAACCTCA TGGCTCACCT TCAATATCAA TAAATTCCTC  
 901 CTCACAAATA TGAAGCTTAG AGAAGCCTTA GCATCAGCCT TAGATAAGGA  
 951 AGCTCTTGTC TCAACTATAT TCTTAGGCCG TGCAAAACT GCCGATCATC  
 1001 TCCTACCTAC AATATTCAT AGCTATCCCG AACATCAAAA ACAAGAGATG  
 30 1051 GCACAACGCC AAGCTTACGC TAAAAAATC TTTAAGAAG CTTTAGAAGA  
 1101 AATCCAAATC ACTGCTAAG ATCTCGAACA TCTTAATCTT ATCTTTCCTG  
 1151 TTTCTCTGTC AGCAAGTTCT TTAGTAGTCC AACTTATACG AGAACAGTGG  
 1201 AAAGAAAGTT TAGGGTTCGC TATCCCTATT GTCGGAAGG AATTGCTCT  
 1251 TCTCAAGCA GACCTATCTT CAGGGAACCT CTCCTTAGCT ACAGGAGGAT  
 35 1301 GGTTGCGAGA CTTTGCTGAT CCTATGGCAT TTCTAACGAT CTTTGCTTAT  
 1351 CCATCAGGAG TTCTCTCTTA TGCAATCAAC CATAAGGACT TCCTAGAAAT  
 1401 TCTACAAAAC ATAGAACAAG AGCAAGATCA CCAAAACGC TCGGAATTAG  
 1451 TGTCGCAAGC TTCTCTTTAC CTAGAGACCT TTCATATTAT TGAGCCGATC  
 1501 TACCACGACG CATTTCAAT TGCTATGAAT AAAAACTTT CTAATCTAGG  
 40 1551 AGTCTACCA ACAGGAGTTG TGGACTTCCG TTATGCTAAG GAAAATTAG

The PSORT algorithm predicts that the protein is an outer membrane lipoprotein (0.790).

The protein was expressed in *E.coli* and purified both as a GST-fusion product and a His-tag fusion product. Purification of the protein as a GST-fusion product is shown in Figure 12A. The recombinant proteins were used to immunise mice, whose sera were used in Western blots (Figures 12B and 12C). FACS analysis was also performed.

These experiments show that cp6466 is a useful immunogen. These properties are not evident from the sequence alone.

### Example 13

The following *C.pneumoniae* protein (PID 4376468) was expressed <SEQ ID 25; cp6468>:

50 1 MFSRWITLFL LFISLTGCS YSSKHQSLI IPIHDDPVAF SPEQAKRAMD  
 51 LSIAQLLFDG LTRETHRESN DLRLAIASRY TVSEDFCSYT FFIKDSALWS  
 101 DGTPITSEDI RNAWEYAQEN SPHIQIFQGL NFSTPSSNAI TIHLDSPNPD  
 151 FPKLLAFPAF AIFKPNPKL FSGPYTLVEY FPGHNIHLKK NPNYYDYHCV  
 201 SINSIKLLII PDIYTAIHL NRKVDWVGQ PWHQGI PWEL HKQSQYHYT  
 55 251 YPVEGAPWLC LNTKSPHLND LQNRHRLATC IDKRSIIEEA LQGTQQPAET

301 LSRGAPQPNQ YKKQKPLTPQ EKLVLTYPSD ILRCQRIAEI LKEQWKAAGI  
 351 DLILEGLEHYH LFNKRRKVD YAIATQTGVA YYPGANLISE EDKLLQNFEI  
 401 IPIYYLSYDY LTQDFIEGVI YNAGAVDLK YTYFP\*

A predicted signal peptide is highlighted.

5 The cp6468 nucleotide sequence <SEQ ID 26> is:

1 ATGTTTTCAC GATGGATCAC CCTCTTTTTA TTATTCATTA GCCTTACTGG  
 51 ATGCTCCTCC TACTCTTCAA AACATAACA ATCTTTAATT ATTCCCATAC  
 101 ATGACGACCC TGTAGCTTTT TCTCTGAAC AAGCAAAACG GGCCATGGAC  
 151 CTTTCTATTG CCCAACTTCT TTTTGATGGT CTGACTAGAG AAACATCATG  
 10 201 CGAATCCAAT GATTTGGAAT TAGCGATTGC CAGTCGCTAT ACAGTCTCTG  
 251 AAGACTTTTG CTCTTATACG TTCCTTATCA AAGACAGCGC TTTATGGAGC  
 301 GACGGAACAC CAATCACCTC CGAAGATATC CGTAACGCTT GGGAGTATGC  
 351 ACAGGAGAAC TCTCCCCACA TACAGATCTT CCAAGGACTT AACTTCTCAA  
 401 CTCTTCATC AAATGCAATT ACGATTATC TCGACTCGCC CAACCCGAT  
 15 451 TTTCTTAAGC TTCCTGCTTT TCCTGCATTT GCTATCTTTA AACCAGAAAA  
 501 CCCGAAGCTC TTTAGCGGTC CGTATACTCT TGTAGAGTAT TTCCCAGGGC  
 551 ATAACATTCA TTTAAAGAAA AACCCTAACT ATTACGACTA CCACTGCGTC  
 601 TCCATCAACT CCATCAAACCT GCTCATTATT CCTGATATAT ATACAGCCAT  
 651 CCACCTCCTA AACAGAGGCA AGGTGGACTG GGTAGACAA CCCTGGCATC  
 20 701 AAGGGATTCC TTGGGAGCTC CATAACAAT CGCAATATCA CTACTACACC  
 751 TATCCTGTAG AAGGTGCCTT CTGGCTTTGT CTAATACAA AATCCCCACA  
 801 CTTAAATGAT CTTCAAAACA GACATAGACT CGCTACTTGT ATTGATAAAC  
 851 GTTCTATCAT TGAAGAAGCT CTTCAAGGAA CCCAACAACC AGCGGAAACA  
 901 CTGTCCCGAG GAGCTCCACA ACCAAATCAA TATAAAAAAC AAAAGCCTCT  
 25 951 AACTCCACAA GAAAACTCG TGCTTACCTA TCCCTCAGAT ATTCTAAGAT  
 1001 GCCAACGCAT AGCAGAAATC TTAAGGAAC AATGGAAAGC TGCTGGAATA  
 1051 GATTTAATCC TTGAAGGACT CGAATACCAT CTGTTTGTTA ACAAACGAAA  
 1101 AGTCCAAGAC TACGCCATAG CAACACAGAC TGGAGTTGCT TATTACCCAG  
 1151 GAGCAAATCT AATTTCTGAA GAAGACAAGC TCCTGCAAAA CTTTGAGATT  
 30 1201 ATCCCGATCT ACTATCTGAG CTATGACTAT CTCACCTCAAG ATTTTATAGA  
 1251 GGGAGTAATC TATAATGCTT CTGGAGCTGT AGATCTCAA TATACCTATT  
 1301 TCCCCTAG

The PSORT algorithm predicts that this protein is an outer membrane lipoprotein (0.790).

The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 13A.

35 The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 13B) and for FACS analysis. A his-tagged protein was also expressed.

These experiments show that cp6468 is a useful immunogen. These properties are not evident from the sequence alone.

#### Example 14

40 The following *C.pneumoniae* protein (PID 4376469) was expressed <SEQ ID 27; cp6469>:

1 MMHRLKPTL KSLIPNLLFL LLTLSSCSKQ KQEPLGKHLV IAMSHDLADL  
 51 DPRNAYLSRD ASLAKALYEG LTRETDQGIA LALAESYTLS KDHKVYTFKL  
 101 RPSVWSDGTP LTAYDFEKSI KQLYFEEFSP SIHTLLGVIK NSSAIHNAQK  
 151 SLETLGIAQK DDLTLVITL QPPFYFLTLI ARPVFSVHH TLRESYKKG  
 45 201 PPSTYISNGP FVLKKHEHQN YLILEKNPHY YDHESVKLDR VTLKIIPDAS  
 251 TATKLFKSKS IDWIGSPWSA PISNEDQKVL SQEKILTVSV SSTLLIYNL  
 301 QKPLIQNKAL RKALAHADR KSILRLVPSG QEAVTLVPPN LSQNLQKEI  
 351 STEERQTKAR AYFQEAETL SEKELAELSI LYPIDSSNS IIAQEIQRQL  
 401 KDTLGLKIKI QGMEYHCFLK KRRQGDFFIA TGGWIAEYVS PVAFLSILGN  
 50 451 PRDLTQWRNS DYKFTLEKLY LPHAYKENLK RAEMIIEET PIIPLYHGKY  
 501 IYAIHPKIQN TFGSLGHTD LKNIDILS\*

A predicted signal peptide is highlighted.

The cp6469 nucleotide sequence <SEQ ID 28> is:

```

1  ATGAAGATGC ATAGGCTTAA ACCTACCTTA AAAAGTCTGA TCCCTAATCT
51  TCTTTTCTTA TTGCTCACTC TTTCAGCTG CTCAAAGCAA AAACAAGAAC
101 CCTTAGGAAA ACATCTCGTT ATTGCGATGA GCCATGATCT CGCCGACCTA
151 GATCCTCGCA ATGCCATATT AAGCAGAGAT GCTTCCCTAG CAAAAGCCCT
5  201 CTATGAAGGA CTGACAAGAG AAACGTATCA AGGAATCGCA CTGGCTCTTG
251 CAGAAAGTTA TACCTGTGCA AAAGATCATA AGGTCTATAC CTTTAAACTC
301 AGACCTTCTG TGTGGAGCGA TGGCACTCCA CTCCTGCTT ATGACTTTGA
351 AAAATCTATA AAACAACGTG ACTTCGAAGA ATTTTCACCT TCCATACATA
10 401 CTTTACTCGG CGTGATTAAT AATTCTTCGG CAATCCACAA TGCTCAAAAA
451 TCTCTGGAAT CTCTTGGGAT ACAGGCAAAA GATGATCTTA CTTTGGTGAT
501 TACCTTAGAG CAACCTTTCC CATACTTTCT CACACTTATC GCTCGCCCCG
551 TATTCTCCCC TGTTCATCAC ACCCTTAGGG AATCCTATAA GAAAGGAACA
601 CCCCCTATCCA CATACTCTC CAATGGGCC TTTGTCTTAA AAAACATGA
15 651 ACACCAAAAC TACTTAATTT TAGAAAAAAA TCCTCACTAC TATGATCATG
701 AATCAGTAAA GTTAGACCGA GTCACCTTAA AAATTATCCC AGACGCCTCC
751 ACAGCCACGA AACTTTTCAA AAGTAAATCT ATAGATTGGA TTGGCTCACC
801 TTGGAGCGCT CCGATATCTA ACGAAGACCA AAAAGTTCTC TCCCAAGAAA
851 AGATTCTTAC TATTCTGTG TCAAGCACC CCCTTCTTAT CTATAACCTG
20 901 CAAAAACCTC TAATACAAAA TAAAGCCCTC AGGAAAGCCA TTGCTCATGC
951 TATTGATAGA AAATCTATCT TAAGACTCGT GCCTTCAGGA CAAGAAGCTG
1001 TAACTCTAGT TCCCCCAAAT CTTTCACAAC TCAATCTTCA AAAAGAGATC
1051 TCAACAGAAAG AACGACAAAC AAAAGCCAGA GCATATTTTC AAGAAGCTAA
1101 AGAAACACTT TCTGAAAAAG AACTCGCAGA ACTCAGCATC CTCTATCCTA
25 1151 TAGATTCTCT GAATTCCTCC ATCATAGCTC AAGAAATCCA AAGACAACCT
1201 AAAGATACCT TAGGATTGAA AATCAAAATC CAAGGCATGG AGTACCACTG
1251 CTTTTTAAAG AAACGTCGTC AAGGAGATT TTTTCATAGC ACAGGAGGAT
1301 GGATTGCGGA ATACGTAAGC CCCGTAGCCT TCCTATCTAT TCTAGGCAAC
1351 CCCAGAGACC TCACACAATG GAGAAACAGT GATTACGAAA AGACTTTAGA
1401 GAAACTCTAT CTCCCTCATG CCTACAAAGA GAATTTAAAA CGCGCAGAAA
30 1451 TGATAATAGA AGAAGAAACC CCGATTATCC CCCTGTATCA CGGCAATAT
1501 ATTTACGCTA TACATCTTAA AATCCAGAA ACATTGCGAT CTCTTCTAGG
1551 CCACACAGAT CTCAAAAATA TCGATATCTT AAGTTAG

```

The PSORT algorithm predicts a periplasmic location (0.934).

The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 14A.

35 The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 14B) and for FACS analysis. A his-tagged protein was also expressed.

These experiments show that cp6469 is a useful immunogen. These properties are not evident from the sequence alone.

#### Example 15

40 The following *C.pneumoniae* protein (PID 4376602) was expressed <SEQ ID 29; cp6602>:

```

1  MAASGGTGGL GGTQGVNLAA VEAAAAKADA AEVVASQEGS EMNMIQQSQD
51  LTNPAATRT KKKKEKFQTL ESRKKGEAGK AEKKSESTEE KPDTDLADKY
101 ASGNSEISGQ ELRGLRDAIG DDASPEDILA LVQEKIKDPA LQSTALDYLV
151 QTTPPSQGKL KEALIQARNT HTEQFGRTAI GAKNILFASQ EYADQLNVSP
45 201 SGLRSLYLEV TGDTHTCQDL LSMLQDRYTY QDMAIVSSFL MKGMATELKR
251 QGPYVPSAQL QVLMTEIRNL QAVLTSYDYF ESRVPILLDS LKAEGIQTPS
301 DLNFVKVAES YHKIINDKFP TASKVEREVR NLIGDDVDSV TGVNLNFFSA
351 LRQTSRLFS SADKRQQLGA MIANALDAVN INNEDYPKAS DFPKPYPWS*

```

The cp6602 nucleotide sequence <SEQ ID 30> is:

```

50 1  ATGGCAGCAT CAGGAGGCAC AGGTGGTTTA GGAGGCACTC AGGGTGTCAA
51  CCTTGAGCT GTAGAAGCTG CAGCTGCAAA AGCAGATGCA GCAGAAGTTG
101 TAGCCAGCCA AGAAGGTTCT GAGATGAACA TGATTCAACA ATCTCAGGAC
151 CTGACAAATC CCGCAGCAGC AACACGCACG AAAAAAAGG AAGAGAAGTT
201 TCAAACTCTA GAATCTCGGA AAAAAGGAGA AGCTGGAAGG GCTGAGAAAA
55 251 AATCTGAATC TACAGAAGAG AAGCCTGACA CAGATCTTGC TGATAAGTAT
301 GCTTCTGGGA ATTCTGAAAT CTCTGGTCAA GAACCTCGCG GCCTGCGTGA
351 TGCAATAGGA GACGATGCTT CTCCAGAAGA CATTCTTGCT CTTGTACAAG

```

-55-

401 AGAAAATTAA AGACCCAGCT CTGCAATCCA CAGCTTTGGA CTACCTGGTT  
 451 CAAACGACTC CACCCCTCCCA AGGTAAATTA AAAGAAGCGC TTATCCAAGC  
 501 AAGGAATACT CACACGGAGC AATTCGGACG AACTGCTATT GGTGCGAAAA  
 551 ACATCTTATT TGCCTCTCAA GAATATGCAG ACCAACTGAA TGTTTCTCCT  
 5 TCAGGGCTTC GCTCTTTGTA CTTAGAAGTG ACTGGAGACA CACATACCTG  
 601 TGATCAGCTA CTTTCTATGC TTCAAGACCG CTATACCTAC CAAGATATGG  
 651 CTATTTGTGAG CTCCTTTTCTA ATGAAAGGAA TGGCAACAGA ATTA AAAAGG  
 701 CAGGGTCCCT ACGTACCCAG TCGGCAACTA CAAGTTCTCA TGACAGAAAC  
 751 TCGTAACCTG CAAGCAGTTC TTACCTCGTA CGATTACTTT GAAAGTCGCG  
 801 TTCTTATTTT ACTCGATAGC TTA AAAAGCTG AGGGAATCCA AACTCCTTCT  
 851 GATCTAAACT TTGTGAAGGT AGCTGAGTCC TACCATAAAA TCATTAAACGA  
 901 TAAGTCCCA ACAGCATCTA AAGTAGAACG AGAAGTCCGC AATCTCATAG  
 951 GAGACGATGT TGATTCTGTG ACCGGTGTCT TGA ACTTATT CTTTCTGTCT  
 1001 TTACGTCAAA CGTCGTCACG CCTTTTCTCT TCAGCAGACA AACGTCAGCA  
 1051 ATTAGGAGCT ATGATTGCTA ATGCTTTAGA TGCTGTAAAT ATAAACAATG  
 1101 AAGATTATCC CAAAGCATCA GACTTCCCTA AACCTATATC TTGGTCATGA  
 1151

The PSORT algorithm predicts a cytoplasmic location (0.080).

The protein was expressed in *E.coli* and purified as both a His-tag and a GST-fusion product, as shown in Figure 15A. The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (Figure 15B) and for FACS analysis (Figure 15C).

The cp6602 protein was also identified in the 2D-PAGE experiment (Cpn0324).

These experiments show that cp6602 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

#### Example 16

The following *C.pneumoniae* protein (PID 4376727) was expressed <SEQ ID 31; cp6727>:

1 MKYSLPWLIT SSALVFSLHP LMAANTDLSS SDNYENGSSG SAAFTAKETS  
 51 DASGTYTTLT SDVSIITNVSA ITPADKSCFT NTGGALSFVG ADHSLVLQTI  
 101 ALTHDGAAIN NTNTALSFSF FSSLLIDSAP ATGTSGGKGA ICVTNTEGGT  
 151 ATFTDNASVT LQKNTSEKDG AAVSAYSIDL AKTTTAALLD QNTSTKNGGA  
 201 LCSTANTTVQ GNSGTVTFSS NTATDRGGGI YSKEKDSLTD ANTGVVTFKS  
 251 NTAKTGGAWS SDDNLALTGN TQVLFQENKT TGSAAQANNP EGCGGAICCY  
 301 LATATDKTGL AISQNQEMSF TSNTTTANGG AIYATKCTLD GNTTLTFDQN  
 351 TATAGCGGAI YTETEDFSLK GSTGTVTFST NTAKTGGALY SKGNSSLTGN  
 401 TNLLFSGNKA TGPSNSSANQ EGCGGAILAF IDSGSVSDKT GLSIANNQEV  
 451 SLTSNAATVS GGAIYATKCT LTGNGSLTFD GNTAGTSGGA IYTETEDFTL  
 501 TGSTGTVTF S TNTAKTGGAL YSKGNSLSG NTNLLFSGNK ATGPSNSSAN  
 551 QEGCGGAILS FLESASVSTK KGLWIEDNEN VLSGNTATV SGGAIYATKC  
 601 ALHGNTTLTF DGNTAETAGG AIYTETEDFT LTGSTGTVTF STNTAKTAGA  
 651 LHKTGNTSFT KNKALVPSGN SATATATTTT DQEGCGGAIL CNISESDIAT  
 701 KSLTLTENES LSFINTAKR SGGGIYAPKC VISGSESINF DGNTAETS GG  
 751 AIYSKNLSIT ANGPVSFTNN SGGKGGAIYI ADSGELSLEA IDGDTFSGN  
 801 RATEGTSTPN SIHLGAGAKI TKLAAAPGHT IYFYDPITME APASGGTIEE  
 851 LVINPVVKAI VPPPQPKNGP IASVPVVPVA PANPNTGTIV FSSGKLPSQD  
 901 ASIPANTTTI LNQKINLAGG NVVLKEGATL QVYSFTQOPD STVFMDAGTT  
 951 LETTTTNNFD GSIDLKNSLV NLDALDGKRM ITIAVNSTSG GLKISGDLKF  
 1001 HNEGFSFYDN PGLKANLNL PFLDLSSTSGT VNLDDFNPIP SSMAAPDYGY  
 1051 QGSWTLVPKV GAGGRVTLVA EWQALGYTPK PELRATLVPN SLWNAYVNIH  
 1101 SIQQELATAM SDAPSHPGIW IGGIGNAFHQ DKQKENAGFR LISRGYIVGG  
 1151 SMTTPQEYTF AVAFSQLPGK SKDYVVS DIK SQVYAGSLCA QSSYVIPLHS  
 50 SLRRHVLSKV LPELPGETPL VILHGQVSYGR NHHNMTTKLA NNTQGRSDWD  
 1251 SHSFAVEVGG SLPVDLNYRY LTSYSPYVKL QVVSVNQKGF QEVAADPRIF  
 1301 DASHLVNYSI PMGLTFKHES AKPPSALLLT LGYAVDAYRD HPHCLTSLTN  
 1351 GTSWSTFATN LSRQAFFAEA SGHLKLLHGL DCFASGSCEL RSSRSRYNAN  
 1401 CGTRYSF\*

A predicted signal peptide is highlighted.

The cp6727 nucleotide sequence <SEQ ID 32> is:

```

1  ATGAAATATT CTTTACCTTG GCTACTTACC TCTTCGGCTT TAGTTTTCCTC
51 CCTACATCCA CTAATGGCTG CTAACACGGA TCTCTCATCA TCCGATAACT
101 ATGAAAATGG TAGTAGTGGT AGCGCAGCAT TCACTGCCAA GGAAACTTCG
5   GATGCTTCAG GAACTACCTA CACTCTCACT AGCGATGTTT CTATTACGAA
201 TGTATCTGCA ATTACTCCTG CAGATAAAAG CTGTTTACCA AACACAGGAG
251 GAGCATTGAG TTTTGTGTGA GCTGATCACT CATTGGTTCT GCAAACCATA
301 GCGCTTACGC ATGATGGTGC TGCAATTAAC AATACCAACA CAGCTCTTTC
351 CGACAGCAGG TTCTCGTCAC TCTTAATCGA CTCAGCTCCA GCAACAGGAA
10  CTTTCGGGCGG CAAGGGTGCT ATTTGTGTGA CAAATACAGA GGGAGTACT
451 GCGACTTTTA CTGACAATGC CAGTGTCACT CTCCAAAAAA ATACTTCAGA
501 AAAAGATGGA GCTGCAGTTT CTGCCATACAG CATCGATCTT GCTAAGACTA
551 ACCACAGCAGC TCTCTTAGAT CAAAATACTA GCACAAAAAA TGGCGGGGCC
601 CTCTGTAGTA CAGCAAAACAC TACAGTCCAA GGAAACTCAG GAACGGTGAC
15  CTTCTCCTCA AATACTGCTA CAGATAAAGG TGGGGGGATC TACTCAAAAG
701 AAAAGGATAG CACGCTAGAT GCCAATACAG GAGTCGTTAC CTTCAAATCT
751 AATACTGCAA AGACGGGGGG TGCTTGGAGC TCTGATGACA ATCTTGCTCT
801 TACCGGCAAC ACTCAAGTAC TTTTTCAGGA AAATAAAACA ACCGGCTCAG
851 CAGCACAGGC AAATAACCCG GAAGGTGTG GTGGGGCAAT CTGTTGTTAT
20  CTTGCTACAG CAACAGACAA AACTGGATTA GCCATTTCTC AGAATCAAGA
951 AATGAGCTTC ACTAGTAATA CAACAACCTG GAATGGTGA GCGATCTACG
1001 CTACTTAAATG TACTCTGGAT GGAAACACAA CTCTTACCTT CGATCAGAAT
1051 ACTGCGACAG CAGGATGTGG CGGAGCTATC TATACAGAAA CTGAAGATTT
1101 TTCTCTTAAG GGAAGTACGG GAACCGTGAC CTTTCAGCACA AATACAGCAA
25  AGACAGGCGG CGCCTTATAT TCTAAAGGAA ACAGCTCGCT GACTGGAAAT
1201 ACCAACCTGC TCTTTTCAGG GAACAAAGCT ACGGGCCCGA GTAATTCCTT
1251 AGCAAATCAA GAGGGTTGCG GTGGGGCAAT CCTAGCCTTT ATTGATTCAG
1301 GATCCGTAAG CGATAAAACA GGACTATCGA TTGCAAAACA CCAAGAAGTC
1351 AGCCTCACTA GTAAATGCTG AACAGTAAGT GGTGTGCGA TCTATGCTAC
30  CAAATGTACT CTAATGGAA ACGGCTCCCT GACCTTTGAC GGCAATACTG
1401 CTGGAACCTC AGGAGGGGCG ATCTATACAG AAACGGAAGA TTTTACTCTT
1501 ACAGGAAGTA CAGGAACCGT GACCTTCAGC ACAAATACAG CAAAGACAGG
1551 CGGCCTCTTA TATTCTAAAG GCAACAACCT TCTGTCTGGT AATACCAACC
1601 TGCTCTTTTC AGGGAACAAA GCTACGGGCC CGAGTAATTC TTCAGCAAT
35  CAAGAGGGTT GCGGTGGGGC AATCCTATCG TTTCTTGAGT CAGCATCTGT
1701 AAGTACTAAA AAAGGACTCT GGATTGAAGA TAACGAAAC GTGAGTCTCT
1751 CTGTAAATAC TGCAACAGTA AGTGGCGGTG CGATCTATGC GACCAAGTGT
1801 GCTCTGCATG GAAACACGAC TCTTACCTTT GATGGCAATA CTGCCGAAAC
40  TGCAGGAGGA GCGATCTATA CAGAAACCGA AGATTTTACT CTTACGGGAA
1901 GTACGGGAAC CGTGACCTTC AGCACAAATA CAGCAAAGAC AGCAGGGGCT
1951 CTACATACTA AAGGAAATAC TTCCTTTACC AAAAAAAGG CTCTTGTTAT
2001 TTCTGGAAAT TCAGCAACAG CAACAGCAAC AACAACCTACA GATCAAGAAG
2051 TTGTGTTGG AGCGATCCTC TGTAATATCT CAGAGTCTGA CATAGCTACA
2101 AAAAGCTTAA CTCTTACTGA AAATGAGAGT TTAAGTTTCA TTAACAATAC
45  GGCAAAAAGA AGTGGTGGTG GTATTTATGC TCCTAAGTGT GTAATCTCAG
2201 GCAGTGAATC CATAAATTTT GATGGCAATA CTGCTGAAAC TTCGGGAGGA
2251 GCGATTTATT CGAAAAACCT TTCGATTACA GCTAACGGTC CTGCTCTCCT
2301 TACCAATAAT TCTGGAGGCA AGGGAGGCGC CATTTATATA GCCGATAGCG
2351 GAGAACTTTC CTTAGAGGCT ATTGATGGGG ATATTACTTT CTCAGGGAAC
50  CGAGCGACTG AGGGAACCTT AACTCCCAAC TCGATCCATT TAGGTGCAGG
2451 GGCTAAGATC ACTAAGCTTG CAGCAGCTCC TGGTCATACG ATTTATTTTT
2501 ATGATCCTAT TACGATGGAA GCTCCTGCAT CTGGAGGAAC AATAGAGGAG
2551 TTAGTCATCA ATCCTGTTGT CAAAGCTATT GTTCTCTCTC CCCAACCAAA
55  AAATGGTCCT ATAGCTTCAG TGCCTGTAGT CCCTGTAGCA CCTGCAAACC
2651 CAAACACGGG AACTATAGTA TTTTCTTCTG GAAAACTCCC CAGTCAAGAT
2701 GCCTCGATTG CTGCAAAATC TACCACCATA CTGAACCAGA AGATCAACTT
2751 AGCAGGAGGA AATGTCGTTT TAAAAGAAGG AGCCACCTTA CAAGTATATT
2801 CCTTCACACA GCAGCCTGAT TCTACAGTAT TCATGGATGC AGGAACGACC
60  TTAGAGACCA CGACAACATA CAATACAGAT GGCAGCATCG ATCTAAAGAA
2901 TCTCTCTGTA AATCTGGATG CTTTAGATGG CAAGCGTATG ATAACGATG
2951 CCGTAAACAG CACAAGTGGG GGATTAAAAA TCTCAGGGGA TCTGAAATTC
3001 CATAACAATG AAGGAAGTTT CTATGACAAT CCTGGGTGTA AAGCAAACCT
3051 AAATCTTCCT TTCTTAGATC TTTCTTCTAC TTCAGGAACT GTAAATTTAG
3101 ACGACTTCAA TCCGATTCCT TCTAGCATGG CTGCTCCGGA TTATGGGTAT
65  CAAGGGAGTT GGACTCTGGT TCCTAAAGTA GGAGCTGGAG GGAAGGTGAC
3201 TTTGGTCCGG GAATGGCAAG CGTTAGGATA CACTCTTAAA CCAGAGCTTC
3251 GTGCGACTTT AGTTCCTAAT AGCCTTTGGA ATGCTTATGT AAACATCCAT

```

-57-

5  
 10  
 15  
 20

```

3301 TCTATACAGC AGGAGATCGC CACTGCGATG TCGGACGCTC CCTCACATCC
3351 AGGGATTTGG ATTGGAGGTA TTGGCAACGC CTTCCATCAA GACAAGCAAA
3401 AGGAAAATGC AGGATTCCTG TTGATTTCCA GAGGTTATAT TGTTGGTGGC
3451 AGCATGACCA CCCCTCAAGA ATATACCTTT GCTGTTGCAT TCAGCCAAC
3501 CTTTGGCAAA TCTAAGGATT ACGTAGTCTC GGATATTAAA TCTCAAGTCT
3551 ATGCAGGATC TCTCTGTGCT CAGAGCTCTT ATGTCATTCC CCTGCATAGC
3601 TCATTACGTC GCCACGTCCT CTCTAAGGTC CTTCCAGAGC TCCCAGGAGA
3651 AACTCCCTTT GTTCTCCATG GTCAAGTTTC CTATGGAAGA AACCACCATA
3701 ATATGACGAC AAAGCTTGCG AACAAACAC AAGGGAATC AGACTGGGAC
3751 AGCCATAGCT TCGCTGTTGA AGTCGGTGGT TCTCTTCCTG TAGATCTAAA
3801 CTACAGATAC CTTACCAGCT ACTCTCCCTA TGTGAAACTC CAAGTTGTGA
3851 GTGTAAATCA AAAAGGATTC CAAGAGGTTG CTGCTGATCC ACGTATCTTT
3901 GACGCTAGCC ATCTGGTCAA CGTGTCTATC CCTATGGGAC TCACCTTCAA
3951 ACACGAATCA GCAAAGCCCC CCAGTGCTTT GCTTCTTACT TTAGGTTACG
4001 CTGTAGATGC TTACCGGGAT CACCCTCACT GCCTGACCTC CTTAACAAAT
4051 GGCACCTCGT GGTCTACGTT TGCTACAAAC TTATCACGAC AAGCTTTCTT
4101 TGCTGAGGCT TCTGGACATC TGAAGTTACT TCATGGTCTT GACTGCTTCG
4151 CTTCTGGAAG TTGTGAAGTG CGCAGCTCCT CAAGAAGCTA TAATGCAAA
4201 TGTGGAATC GTTATTCTTT CTA
  
```

20 The PSORT algorithm predicts an outer membrane location (0.915).

The protein was expressed in *E.coli* and purified as a his-tag product, as shown in Figure 16A. The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 16B) and for FACS analysis (Figure 16C). A GST-fusion protein was also expressed.

The cp6727 protein was also identified in the 2D-PAGE experiment (Cpn0444).

25 These experiments show that cp6727 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 17

The following *C.pneumoniae* protein (PID 4376731) was expressed <SEQ ID 33; cp6731>:

30  
 35  
 40  
 45

```

1  MKSSLHWFLI SSSLALPLSL NFSAFAAVVE INLGPTNSFS GPGTYTPPAQ
51  TTNADGTIYN LTGDVSIITNA GSPTALTASC FKETTGNLSF QGHGYQFLQ
101 NIDAGANCTF TNTAANKLLS FSGPSYLSLI QTTNATTGTG AIKSTGACSI
151 QSNYSCYFGQ NFSNDNGGAL QGSSISLSLN PNLTFANKKA TQKGGALYST
201 GGITINNTLN SASFSENTAA NNGGAIYTEA SSFISSNKAI SFINNSVTAT
251 SATGAIYCS STSAPKPVLT LSDNGELNFI GNTAITSGBA IYTDNLVLS
301 GGPTLFPKNS AIDTAAPLGG AIAIADSGSL SLSALGGDIT FEGNTVVKGA
351 SSSQTPTRNS INIGNTNAKI VQLRASQNT IYFYDPITTS ITAALSDALN
401 LINGPDLAGN AYQGTIVFSG EKLSEAEAAE ADNLKSTIQQ PLTLAGGQLS
451 LKSGVTLVAK SFSQSPGSTL LMDAGTTLET ADGITINNLV LNVDLSKETK
501 KATLKATQAS QTVTLSGSL LVDPSGNVYE DVSWNNPQVF SCLTLTADDP
551 ANIHITDLAA DPLEKNPIHW GYQGNWALSW QEDTATKSKA ATLTWTKTGY
601 NPNPERRGTL VANTLWGSFV DVRSIQQLVA TKVRQSQETR GIWCEGISNF
651 PHKDSKINK GFRHISAGYV VGATTTLASD NLITAAFCQL FGKDRDHFIN
701 KNRASAYAAS LHLQHLATLS SPSLLRYLPG SBSEQPVLPD AQISYIYSKN
751 TMKTYTQAP KGESSWYNDG CALELASSLP HTALSHEGLF HAYFPFIKVE
801 ASYIHQDSFK ERNTTLVRSF DSGDLINVSF PIGITFERFS RNERASYEAT
851 VIYVADVYRK NPDCTTALLI NNTSWKTGT NLSRQAGIGR AGIFYAFSPN
901 LEVTSNLSME IRGSSRSYNA DLGGKFPQF*
  
```

A predicted signal peptide is highlighted.

The cp6731 nucleotide sequence <SEQ ID 34> is:

50

```

1  ATGAAATCCT CTCTTCATTG GTTTTAAATC TCGTCATCTT TAGCACTTCC
51  CTTGTCACTA AATTTCTCTG CGTTTGCTGC TGTGTTGAA ATCAATCTAG
101 GACCTACCAA TAGCTTCTCT GGACCAGGAA CCTACACTCC TCCAGCCCAA
151 ACAACAAATG CAGATGGAAC TATCTATAAT CTAACAGGGG ATGCTCTCAAT
201 CACCAATGCA GGATCTCCGA CAGCTCTAAC CGCTTCTGCT TTAAAGAAA
  
```



251 CTACTGGGAA TCTTTCTTTC CAAGGCCACG GCTACCAATT TCTCCTACAA  
 301 AATATCGATG CGGGAGCGAA CTGTACCTTT ACCAATACAG CTGCAAATAA  
 351 GCTTCTCTCC TTTTCAGGAT TCTCCTATTT GTCACATAA CAAACCACGA  
 401 ATGCTTACCAC AGGAACAGGA GCCATCAAGT CCACAGGAGC TTGTCTTATT  
 451 CAGTCGAACT ATAGTTGCTA CTTTGGCCAA AACTTTTCTA ATGACAATGG  
 501 AGGCGCCCTC CAAGGCAGCT CTATCAGTCT ATCGCTAAAC CCCAACCTAA  
 551 CGTTTGCCAA AAACAAAGCA ACGCAAAAAG GGGGTGCCCT CTATTCCACG  
 601 GGAGGGATTA CAATTAACAA TACGTTAAAC TCAGCATCAT TTTCTGAAAA  
 651 TACCGCGGCG AACAAATGGCG GAGCCATTTA CACGGAAGCT AGCAGTTTTA  
 701 TTAGCAGCAA CAAAGCAATT AGCTTTATAA ACAATAGTGT GACCGCAACC  
 751 TCAGCTACAG GGGGAGCCAT TTACTGTAGT AGTACATCAG CCCCCAAACC  
 801 AGCTTTAACT CTATCAGACA ACGGGGAAC TGAACCTTATA GGAAATACAG  
 851 CAATTTACTAG TGGTGGGGCG ATTTATACTG ACAATCTAGT TCTTTCTTCT  
 901 GGAGGACCTA CGCTTTTAA AAACAACCTCT GCTATAGATA CTGCAGCTCC  
 951 CTTAGGAGGA GCAATTGCCA TTGCTGACTC TGGATCTTTG AGTCTTTCGG  
 1001 CTCTTGGTGG AGACATCACT TTTGAAGGAA ACACAGTAGT CAAAGGAGCT  
 1051 TCTTTCGAGTC AGACCACTAC CAGAAATTCT ATTAACATCG GAAACACCAA  
 1101 TGCTAAGATT GTACAGCTGC GAGCCTCTCA AGGCAATACT ATCTACTTCT  
 1151 ATGATCCTAT AACAACTAGC ATCACTGCAG CTCTCTCAGA TGCTCTAAAC  
 1201 TTAAATGGTC CTGACCTTGC AGGGAATCCT GCATATCAAG GAACCATCGT  
 1251 ATTTTCTGGA GAGAAGCTCT CGGAAGCAGA AGCTGCAGAA CTGATAATC  
 1301 TCAAATCTAC AATTTCAGCAA CCTCTAACTC TTGCGGGAGG GCAACTCTCT  
 1351 CTTAAATCAG GAGTCACTCT AGTTGCTAAG TCCTTTTCGC AATCTCCGGG  
 1401 CTCTACCTCT CTCATGGATG CAGGGACCAC ATTAGAAACC GCTGATGGGA  
 1451 TCACATCAAA TAATCTTGT CTCAATGTAG ATTCTTAAA AGAGACCAAG  
 1501 AAGGCTACGC TAAAAGCAAC ACAAGCAAGT CAGACAGTCA CTTTATCTGG  
 1551 ATCGCTCTCT CTGTAGATC CTTCTGGAAA TGTCTACGAA GATGTCTCTT  
 1601 GGAATAACCC TCAAGTCTTT TCTTGTCTCA CTCCTACTGC TGACGACCCC  
 1651 GCGAATATTC ACATCACAGA CTTAGCTGCT GATCCCCTAG AAAAAAATCC  
 1701 TATCCATTGG GGATACCAAG GGAATTGGGC ATTATCTTGG CAAGAGGATA  
 1751 CTGCGACTAA ATCCAAAGCA GCGACTCTTA CCTGGACAAA AACAGGATAC  
 1801 AATCCGAATC CTGAGCGTCG TGGAACCTTA GTTGCTAACA CGCTATGGGG  
 1851 ATCCTTTGTT GATGTGCGCT CCATACAACA GCTTGTAGCC ACTAAAGTAC  
 1901 GCCAATCTCA AGAACTCGC GGCATCTGGT GTGAAGGGAT CTCGAACCTC  
 1951 TTCCATAAAG ATAGCACGAA GATAAATAAA GGTTTTCGCC ACATAAGTGC  
 2001 AGGTTATGTT GTAGGAGCGA CTACAACATT AGCTTCTGAT AATCTTATCA  
 2051 CTGCAGCCTT CTGCCAATTA TTCGGGAAAG ATAGAGATCA CTTTATAAAT  
 2101 AAAAATAGAG CTTCTGCCTA TGCAGCTTCT CTCCATCTCC AGCATCTAGC  
 2151 GACCTTGTCT TCTCCAAGCT TGTACGCTA CCTTCTGGA TCTGAAAGTG  
 2201 AGCAGCCTGT CCTCTTTGAT GCTCAGATCA GCTATATCTA TAGTAAAAAT  
 2251 ACTATGAAAA CCTATTACAC CCAAGCACCA AAGGGAGAGA GCTCGTGGTA  
 2301 TAATGACGGT TGCCTCTTGG AACTTGCAGG CTCCTTACCA CACACTGCTT  
 2351 TAAGCCATGA GGGTCTCTTC CACGCGTATT TTCCCTTCAT CAAAGTAGAA  
 2401 GCTTCGTACA TACACCAAGA TAGCTTCAAA GAACGTAATA CTACCTTGGT  
 2451 ACGATCTTTC GATAGCGGTG ATTTAATTAA CGTCTCTGTG CCTATTGGAA  
 2501 TTACCTTCGA GAGATTCTCG AGAAACGAGC GTGCGTCTTA CGAAGCTACT  
 2551 GTCATCTACG TTGCCGATGT CTATCGTAAG AATCCTGACT GCACGACAGC  
 2601 TCTCTTAATC AACAAATACCT CGTGGAAAAC TACAGGAACG AATCTCTCAA  
 2651 GACAAGCTGG TATCGGAAGA GCAGGGATCT TTTATGCCTT CTCTCCAAAT  
 2701 CTTGAGGTCA CAAGTAACCT ATCTATGGAA ATTCTGGAT CTTACGCGAG  
 2751 CTACAATGCA GATCTTGGAG GTAAGTTCCA GTTCTAA

The PSORT algorithm predicts an outer membrane location (0.926).

The protein was expressed in *E.coli* and purified as a his-tag product, as shown in Figure 17A. A GST-fusion protein was also expressed. The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (Figure 17B; his-tag) and for FACS analysis (Figure 17C; his-tag and GST-fusion).

The GST-fusion protein also showed good cross-reactivity with human sera, including sera from patients with pneumonitis. Less cross-reactivity was seen with the his-fusion.

These experiments show that cp6731 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 18

The following *C.pneumoniae* protein (PID 4376737) was expressed <SEQ ID 35; cp6737>:

```

5      1  MPLSPKSSSF CLLACLCSAS CAPAETRLGG NRVPPITNQG EEILLTSDFV
      51  CSNFLGASFS SSFINSSSNL SLLGKGLSLT PTSCQAPTNS NYALLSAAET
     101  LTFKNFSSIN FTGNQSTGLG GLIYKDIVF QSIKDLIFTT NRVAYSPASV
     151  TTSATPAITT VTTGASALQP TDSLTVENIS QSIKFFGNLA NFGSAISSSP
     201  TAVVVFNNIT ATMSFSHNFT SSGGGVIYGG SLLFFENNSG CIIIFTANSCV
10    251  NSLKGVTSSS GTYALGSGGA ICIPGTGFEL KNNQKCTFS YNGTPNDAGA
     301  IYAETCNIVG NQGALLLDSN TAARNNGAIC AKVLNIQGRG PIEFSRNRAE
     351  KGAIFIGPS VGDPAKQTST LTIASEGDI AFQGNMLNTK PGIRNAITVE
     401  AGGEIVSLSA QGGSRLVIFYD PITHSLPTTS PSNKDITINA NGASGSVVFT
     451  SKGLSSTELL LPANTTTILL GTVKIASGEL KITDNAVNVN LGFATQGSQ
15    501  LTLGSGGTLG LATPTGAPAA VDFTIGKLAF DPFSFLKRFV VSASVNAGTK
     551  NVTLTGALVL DEHDVTDLYD MVSLQTFVAI PIAVFKGATV TKTGFPDGEI
     601  ATPSHYGYQG KWSYTWSRPL LIPAPDGGFP GGPSPSANTL YAVWNSDTLV
     651  RSTYILDPER YGEIVSNSLW ISFLGNQAFS DILQDVLLID HPGLSITAKA
     701  LGAYVEHTPR QGHEGFSGRY GGYQAALSMN YTDHTTLGLS FGQLYKGTNA
20    751  NPYDSRCSEQ MYLLSFFGQF PIVTQKSEAL ISWKAAYGYS KNHLNTTYLR
     801  PDKAPKSGGQ WHNNSYYVLI SAEHPFLNWC LLTRPLAQAW DLSGFISAEF
     851  LGGWQSKFTE TGDLORSFSR GKGYNVSLPI GCSSQWFTPF KKAPSTLTIK
     901  LAYKPDIVRV NPHNIVTVVS NQESTSISGA NLRRHGLFVQ IHDVVDLTED
     951  TQAFNLNYTFD GKNGFTNHRV STGLKSTF*

```

25 A predicted signal peptide is highlighted.

The cp6737 nucleotide sequence <SEQ ID 36> is:

```

      1  ATGCCTCTTT CTTTCAAATC TTCATCTTTT TGTCTACTTG CCTGTTTATG
     51  TAGTGCAAGT TGCGCGTTTG CTGAGACTAG ACTCGGAGGG AACTTTGTTC
    101  CTCCAATTAC GAATCAGGGT GAAGAGATCT TACTCACTTC AGATTTTGTT
30    151  TGTTCAAACT TCTTGGGGGC GAGTTTTTC AGTTCCTTTA TCAATAGTTC
     201  CAGCAATCTC TCCTTATTAG GGAAGGGCCT TTCCCTAACG TTTACCTCTT
     251  GTCAAGCTCC TACAAATAGT AACTATGCGC TACTTTCTGC CGCAGAGACT
     301  CTGACCTTCA AGAATTTTTC TTCTATAAAC TTTACAGGGA ACCAATCGAC
     351  AGGACTTGGC GGCTCATCT ACAGAAAAGA TATTGTTTTC CAATCTATCA
35    401  AAGATTGTAT CTTCACTACG AACCGTGTG CCTATCTCTC AGCATCTGTA
     451  ACTACGTCGG CAACTCCCGC AATCACTACA GTAACACAG GAGCCTCTGC
     501  TCTCCAACCT ACAGACTCAC TCACGTGCGA AAACATATCC CAATCGATCA
     551  AGTTTTTTGG GAACCTTGCC AACTTCGGCT CTGCAATTAG CAGTTCTCCC
     601  ACGGCAGTCG TTAAATTTCAT CAATAACACC GCTACCATGA GCTTCTCCCA
40    651  TAAGCTTTACT TCGTCAGGAG GCGGCGTGAT TTATGAGGGA AGCTCTCTCC
     701  TTTTGTAAAA CAATCTTGGA TGCATCATCT TCACCGCCAA CTCCTGTGTG
     751  AACAGCTTAA AAGGCGTCAC CCCTTCATCA GGAACCTATG CTTTAGGAAG
     801  TGGCGGAGCC ATCTGCATCC CTACGGGAAC TTTGGAATTA AAAACAATC
     851  AGGGGAAGTG CACCTTCTCT TATAATGGTA CACCAAATGA TGCGGTGCG
45    901  ATCTACGCCG AAACCTGCAA CATCGTAGGG AACCAGGGTG CCTGTCTCCT
     951  AGATAGCAAC ACTGCAGCGA GAAATGGCGG AGCCATCTGT GCTAAAGTGC
    1001  TCAATATTCA AGGACCGGGT CCTATTGAAT TCTCTAGAAA CCGCGCGGAG
    1051  AAGGGTGGAG CTATTTTCAT AGGCCCTCTT GTTGAGACC CTGCGAAGCA
    1101  AAGATCGACA CTTACGATTT TGGCTTCCGA AGGTGATATT GCGTTCCAAG
50    1151  GAAACATGCT CAATACAAAA CCTGGAATCC GCAATGCCAT CACTGTAGAA
    1201  GCAGGGGGAG AGATTGTGTC TCTATCTGCA CAAGGAGGCT CACGCTTGT
    1251  ATTTTATGAT CCCATTACAC ATAGCCTCCC AACCACAAGT CCGTCTAATA
    1301  AAGACATTAC AATCAACGCT AATGCGGCTT CAGGATCTGT AGTCTTTACA
    1351  AGTAAGGGAC TCTCTCTAC AGAACTCTG TTGCCTGCCA ACACGACAAC
55    1401  TATACTTCTA GGAACAGTCA AGATCGCTAG TGGAGAACTG AAGATTACTG
    1451  ACAATGCGGT TGTCAATGTT CTTGGCTTCG CTACTCAGGG CTCAGTCTAG
    1501  CTTACCCCTG GCTCTGGAGG AACCTTAGGG CTGGCAACAC CCACGGGAGC
    1551  ACCTGCCGCT GTAGACTTTA CGATTGGAAA GTTAGCATTC GATCCTTTTT
    1601  CCTTCTCTAA AAGAGATTTT GTTTCAGCAT CAGTAAATGC AGGCACAAAA
60    1651  AACGTCACTT TAACAGGAGC TCTGTTCTT GATGAACATG ACGTTACAGA

```

1701 TCTTTATGAT ATGGTGTCTAT TACAAACTCC AGTAGCAATT CCTATCGCTG  
 1751 TTTTCAAAGG AGCAACCGTT ACTAAGACAG GATTTCCTGA TGGGGAGATT  
 1801 GCGACTCCAA GCCACTACGG CTACCAAGGA AAGTGGTCCT ACACATGGTC  
 1851 CCGTCCCCTG TTAATTCAG CTCCTGATGG AGGATTTCTT GGAGGTCCCT  
 5 1901 CTCCTAGCGC AAATACTCTC TATGCTGTAT GGAATTCAGA CACTCTCGTG  
 1951 CGTTCTACCT ATATCTTAGA TCCCGAGCGT TACGGAGAAA TTGTCAGCAA  
 2001 CAGCTTATGG ATTTCTTCT TAGGAAATCA GGCATTCTCT GATATTCTCC  
 2051 AAGATGTTCT TTTGATAGAT CATCCCGGGT TGTCCATAAC CGCGAAAGCT  
 2101 TTAGGAGCCT ATGTCGAACA CACACCAAGA CAAGGACATG AGGGCTTTTC  
 10 2151 AGGTGCGTAT GGAGGCTACC AAGCTGCGCT ATCTATGAAC TACACGGACC  
 2201 ACACACGTT AGGACTTTCT TTCGGGCAGC TTTATGGAAA AACTAACGCC  
 2251 AACCCTACG ATTCACGTTG CTCAGAACAA ATGTATTTAC TCTCGTTCTT  
 2301 TGGTCAATTC CCTATCGTGA CTCAAAAGAG CGAGGCCTTA ATTTCTTGGA  
 2351 AAGCAGCTTA TGGTTATTCC AAAAATCACC TAAATACCAC CTACCTCAGA  
 15 2401 CCTGACAAAG CTCAAAATC TCAAGGGCAA TGGCATAACA ATAGTTACTA  
 2451 TGTTCCTTAT TCTGCAGAAC ATCCTTTCTT AAACCTGGTG CTTCCTACAA  
 2501 GACCTCTGGC TCAAGCTTGG GATCTTTCAG GTTTTATTTT CGCAGAATTC  
 2551 CTAGGTGCTT GGCAAAGTAA GTTCACAGAA ACTGGAGATC TGCAACGTAG  
 20 2601 CTTTAGTAGA GGTAAAGGGT ACAATGTTTC CCTACCGATA GGATGTTCTT  
 2651 CTCAATGGTT CACACCATT AAGAAGGCTC CTTCCTACACT GACCATCAAA  
 2701 CTTGCCTACA AGCCTGATAT CTATCGTGTG AACCTCACA ATATTGTGAC  
 2751 TGTCGTCTCA AACCAAGAGA GCACCTCGAT CTCAGGAGCA AATCTACGCC  
 2801 GCCACGGTTT GTTTGTACAA ATCCATGATG TAGTAGATCT CACCGAGGAC  
 2851 ACTCAGGCCT TTCTAAACTA TACCTTTGAC GGGAAAAATG GATTTACAAA  
 25 2901 CCACCGAGTG TCTACAGGAC TAAATCCAC ATTTTAA

The PSORT algorithm predicts an outer membrane location (0.940).

The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 18A. The recombinant protein was used to immunise mice, whose sera were used in an immunoblot analysis blot (Figure 18B) and for FACS analysis (Figure 18C). A his-tagged protein was also expressed.

The cp6737 protein was also identified in the 2D-PAGE experiment (Cpn0454) and showed good cross-reactivity with human sera, including sera from patients with pneumonitis.

These experiments show that cp6737 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### 35 Example 19

The following *C.pneumoniae* protein (PID 4377090) was expressed <SEQ ID 37; cp7090>:

1 MNIHSLWKLC TLALALLALPA CSLSPNYGWE DSCNFTCHHTR RKKPSSFGFV  
 51 PLYTEEDFNP NPTFGGYDSK EEKQYKSSQV AAFRNITFAT DSYTIKGEEN  
 101 LAAILTNLVHY MKKNPKATLY IEGHTDERGA ASYNLALGAR RANAIKEHLR  
 40 151 KQGISADRLS TISYGKEHPL NSGHNELAWQ QNRRTEFKIH AR\*

A predicted signal peptide is highlighted.

The cp7090 nucleotide sequence <SEQ ID 38> is:

1 ATGAATATAC ATTCCCTATG GAAACTTTGT ACTTTATTGG CTTTACTTGC  
 51 ATTGCCAGCA TGAGCCCTTT CCCCTAATTA TGGCTGGGAG GATTCTGTGA  
 45 101 ATACATGCCA TCATACAAGA CGAAAAAGC CTTCTTCTTT TGGCTTTGTT  
 151 CCTCTCTATA CCGAAGAGGA CTTTAACCCT AATTTTACCT TCGGTGAGTA  
 201 TGATTCCAAA GAAGAAAAAC AATACAAATC AAGCCAAGTT GCAGCATTTT  
 251 GTAATATCAC CTTTGCTACA GACAGCTATA CAATTAAAGG TGAAGAGAAC  
 301 CTTGCGATTG TCACGAACCT GGTTCACTAC ATGAAGAAAA ACCCGAAAGC  
 50 351 TACACTGTAC ATTGAAGGGC ATACTGACGA GCGTGGAGCT GCATCCTATA  
 401 ACCTTGCTTT AGGAGCACGA CGAGCCAATG CGATTAAAGA GCATCTCCGA  
 451 AAGCAGGGAA TCTCTGACGA TCGTCTATCT ACTATTCTCT ACGGAAAAA

-61-

501 ACATCCTTTA AATTCGGGAC ACAACGAACT AGCATGGCAA CAAAATCGCC  
 551 GTACAGAGTT TAAGATTCAT GCACGCTAA

The PSORT algorithm predicts an outer membrane location (0.790).

The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 19A.

- 5 A his-tagged protein was also expressed. The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (Figure 19B) and for FACS analysis.

These experiments show that cp7090 is useful immunogen. These properties are not evident from the sequence alone.

### Example 20

- 10 The following *C.pneumoniae* protein (PID 4377091) was expressed <SEQ ID 39; cp7091>:

1 MLRQLCFQVF FFCFASLVYA EELEVVVRSE HITLPIEVSC QTDTKDPKIQ  
 51 KYLSSLTEIF CKDIALGDCL QPTAASKESS SPLAISLRH VPQLSVVLLQ  
 101 SSKTPQTLCS FTISQNLSD RQKIHAADT VHVALTGIPG ISAGKIVPAL  
 151 SSLGKDQKLK QGELWTFDYD GKNLAPLITE CSLSITPKWV GVGSNFPYLY  
 15 VSYKYGVPKI FLGSLENTG KKVLPKGNQ LMPTFSRKK LLAFVADTYG  
 201 NPDLFIQFFS LTSGPMGRPR RLLNENFGTQ GNPSFNPEGS QLVFISNKDG  
 251 RPRLYMSLD PEPQAPRLLT KKYRNSSCPA WSPDGKKIAF CSVIKGVRQI  
 301 CIYDLSSGED YQLTTSPTNK ESPSWAIDSR HLVFSAGNAE ESELYLISLV  
 351 TKKTNKIAIG VGEKRFPSWG AFPQQPIKRT L\*

- 20 A predicted signal peptide is highlighted.

The cp7091 nucleotide sequence <SEQ ID 40> is:

1 ATGTTACGGC AACTATGCTT CCAAGTFTTT TTCTTTTGCT TCGCATCGCT  
 51 AGTCTATGCT GAAGAATTAG AAGTTGTGT CCGTTCCGAA CATATCACGC  
 101 TCCCTATTGA GGTCTCTTGC CAGACCGATA CGAAAGATCC AAAAATACAG  
 25 AAATACCTCA GCTCGCTAAC GGAGATATTT TGCAAGGACA TTGCCCTAGG  
 151 AGATTGTCTA CAACCCACAG CGGCTTCTAA AGAATCGTCA TCTCCTTAG  
 201 CAATATCTTT ACGGTTCAT GTACCTCAGC TATCTGTAGT GCTTTTACAG  
 251 TCTTCAAAAA CTCCTCAAAC CTTATGTTCT TTTACTATTT CTCAAAATCT  
 301 TTCTGTAGAT CGTCAAAAAA TCCATCACGC TGCTGATACA GTTCATTACG  
 351 CCCTCACAGG GATTCTTGGA ATCAGTGCTG GGAAAATTGT TTTTGCTCTA  
 401 AGTTCTTTAG GAAAAGATCA AAAGCTCAAG CAAGGAGAAT TATGGACTAC  
 451 AGATTACGAT GGGAAAAACC TCGCCCTTTT AACCACAGAA TGTTCGCTCT  
 501 CTATAACTCC AAAATGGGTG GGTGTGGGAT CAAATTTTCC CTATCTCTAT  
 551 GTTTCGTATA AGTATGGTGT GCCTAAAAAT TTTCTTGGTT CCCTAGAGAA  
 601 CACTGAAGGT AAAAAAGTCC TTCCGTTAAA AGGCAACCAA CTCATGCCTA  
 35 651 CAGTTTCTCC AAGAAAAAAG CTTTGTAGCT TCGTTGCTGA TACGTATGGA  
 701 AATCCTGATT TATTTATTCA ACCGTTCTCA CTAAC TTCAG GACCTATGGG  
 751 TCGCCACGTC CGCTCCTTA ATGAGAATTT CGGGACTCAA GGAATCCCT  
 801 CCTTCAACCC TGAAGGATCC CAGCTTGCTT TTATATCGAA CAAAGACGGC  
 851 CGTCCGCGTC TTTATATTAT GTCCCTCGAT CCTGAACCCC AAGCACCTCG  
 40 901 CTTGCTGACA AAAAAATACA GAAATAGCAG TTGCCCTGCA TGGTCTCCAG  
 951 ATGGTAAAAA AATAGCCTTC TGCTCTGTAA TTAAAGGGGT GCGACAAATT  
 1001 TGTATTACG ATCTCTCCTC TGGAGAGGAT TACCAACTCA CTACGTCCTC  
 1051 CACAAATAAA GAGAGTCCTT CTTGGGCTAT AGACAGCCGT CATCTTGTCT  
 1101 TTAGTGC GGG GAATGCTGAA GAATCAGAGT TATATTTAAT CAGTCTAGTC  
 45 1151 ACCAAAAAAA CTAACAAAT TGCTATAGGA GTAGAGAGAA AACGGTTCCC  
 1201 CTCTTGGGGT GCTTTCCCTC AGCAACCGAT AAAGAGAACA CTATGA  
 1251

The PSORT algorithm predicts an inner membrane location (0.109).

The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 20A.

- 50 A his-tagged protein was also expressed. The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (Figure 20B) and for FACS analysis.

These experiments show that cp7091 is a useful immunogen. These properties are not evident from the sequence alone.

### Example 21

The following *C. pneumoniae* protein (PID 4376260) was expressed <SEQ ID 41; cp6260>:

```

5      1  MRFSLCGFPL VFSFTLLSVF DTSLSATTTIS LTPEDSFHGD SQNAERSYNV
      51  QAGDVYSLTG DVSISNVDNS ALNKACFNVT SGSVTFAGNH HGLYFNMISS
     101  GTTKEGAVLC CQDPQATARF SGFSTLSFTIQ SPGDIKEQGC LYSKNALMLL
     151  NNYVVRFEQN QSKTKGGAIS GANVTIVGNY DSVSFYQNA TFGGAIHSSG
     201  PLQIAVNQAE IRFAQNTAKN GSGGALYSDG DIDIDQNAVY LFRNEALTT
    10  251  AIGKGGAVCC LPTSGSSTPV PIVTFSDNRQ LVFERNHSIM GGGATYARKL
      301  SISSGGPTLF INNISYANSQ NLGGAIAIDT GGEISLSAEK GTITFQGNRT
      351  SLPFLNGIHL LQNAKFLKLQ ARNGYSIEFY DPITSEADGS TQLNINGDPK
      401  NKBYTGTILF SGEKSLANDP RDFKSTIPQN VNLSAGYLV KEGAEVTVSK
      451  FTQSPGSHLV LDLGTKLIAS KEDIAITGLA IDIDSLSSSS TAAVIKANTA
    15  501  NKQISVTD SI ELISPTGNAY EDLRMRNSQT FPLLSLEPGA GGSVTVTAGD
      551  FLFPVSPHYGF QGNWKLAWTG TGNKVGEFFW DKINYKPRPE KEGNLVFNIL
      601  WGNAVDVRS L MQVQETHASS LQTDRLWID GIGNFFHVSA SEDNIRYRHN
      651  SGGYVLSVNN EITPKHYTSM AFSQLFSRDK DYAVSNNNEYR MYLGSYLYQY
      701  TTS LGNIFRY ASRNPVN NVG ILSRRFLQNP LMIFHFLCAY GHATNDMKT D
    20  751  YANFPMVKNS WRNNCWAIEC GGSMPLLVFE NGRLPQGAIP FMKLQLVYAY
      801  QGDFKETAD GRRFSNGSLT SISVPLGIRF EKLALSQDVL YDFSFSYIPD
      851  IFRKDPSC EA ALVISGDSWL VPAAHVSRHA FVSGGTGRYH FNDYTELLCR
      901  GSIECRPHAR NYNINCGSKF RF*

```

A predicted signal peptide is highlighted.

25 The cp6260 nucleotide sequence <SEQ ID 42> is:

```

      1  ATGCGATTMT CGCTCTGCGG ATTTCTCTCTA GTTTTCTCTT TTACATTGCT
     51  CTCAGTCTTC GACACTTCTT TGAGTGCTAC TACGATTTCT TTAACCCCAG
    101  AAGATAGTTT TCATGGAGAT AGTCAGAATG CAGAACGTTC TTATAATGTT
    151  CAAGCTGGGG ATGCTCTATAG CCTTACTGGT GATGTCTCAA TATCTAACGT
    201  CGATAACTCT GCATTAAATA AAGCCTGCTT CAATGTGACC TCAGGAAGTG
    251  TGACGTTTCG AGGAAATCAT CATGGGTAT ATTTTAATAA TATTTCTCTA
    301  GGAACTACAA AGGAAGGGGC TGTACTTTGT TGCCAAGATC CTCAGCAAC
    351  GGCACGTTTT TCTGGGTTCT CCACGCTCTC TTTTATTTCAG AGCCCCGGAG
    401  ATATTAAAGA ACAGGGATGT CTCTATTCAA AAAATGCACT TATGCTCTTA
    451  AACAATTATG TAGTGCGTTT TGAACAAAAC CAAAGTAAGA CTAAAGCGCG
    501  AGCTATTAGT GGGCGGAATG TTACTATAGT AGGCAACTAC GATTCCGTCT
    551  CTTTCTATCA GAATGCAGCC ACTTTTGAG GTGCTATCCA TTCTTCAGGT
    601  CCCCTACAGA TTGCAGTAAA TCAGGCAGAG ATAAGATTGT CACAAAATAC
    651  TGCCAAGAAT GGTCTCGGAG GGGCTTTGTA CTCCGATGGT GATATTGATA
    701  TTGATCAGAA TGCTTATGTT CTATTTCGAG AAAATGAGGC ATTGACTACT
    751  GCTATAGGTA AGGGAGGGGC TGTCTGTTGT CTTCCCACTT CAGGAAGTAG
    801  TACTCCAGTT CCTATTGTGA CTTTCTCTGA CAATAACAG TTAGTCTTTG
    851  AAAGAAACCA TTCCATAATG GGTGGCGGAG CCATTTATGC TAGGAAACTT
    901  AGCATCTCTT CAGGAGGTCC TACTCTATTT ATCAATAATA TATCATATGC
    951  AAAITCGCAA AATTAGGTG GAGCTATTGC CATTGATACT GGAGGGGAGA
   1001  TCAGTTTATC AGCAGAGAAA GGAACAATTA CATTTCAAGG AAACCGGACG
   1051  AGCTTACCGT TTTTGAATGG CATCCATCTT TTACAAAATG CTAAATTCCT
   1101  GAAATTACAG GCGAGAAATG GATACTCTAT AGAATTTTAT GATCCTATTA
   1151  CTTCTGAAGC AGATGGGTCT ACCCAATTGA ATATCAACGG AGATCCTAAA
   1201  AATAAAGAGT ACACAGGGAC CATACTCTTT TCTGGAGAAA AGAGTCTAGC
   1251  AAACGATCCT AGGGATTTTA AATCTACAAT CCCTCAGAAC GTCAACCTGT
   1301  CTGCAGGATA CTTAGTTATT AAAGAGGGGG CCGAAGTCAC AGTTTCAAAA
   1351  TTCACGCAGT CTCCAGGATC GCATTTAGTT TTAGATTTAG GAACCAAACT
   1401  GATAGCCTCT AAGGAAGACA TTGCCATCAC AGGCCTCGCG ATAGATATAG
   1451  ATAGCTTAAG CTCATCTCTA ACAGCAGCTG TTATTAAAGC AAACACCGCA
   1501  AATAAACAGA TATCCGTGAC GGACTCTATA GAACCTATCT CGCCTACTGG
   1551  CAATGCCTAT GAAGATCTCA GAATGAGAAA TTCACAGACG TTCCCTCTGC
   1601  TCTCTTTAGA CCCTGGAGCC GGGGGTAGTG TGACTGTAAC TGCTGGAGAT
   1651  TTCTTACCGG TAAGTCCCCA TTATGGTTTT CAAGGCAATT GGAAATTAGC
   1701  TTGGACAGGA ACTGGAACAA AAGTTGGAGA ATTCTTCTGG GATAAAATAA

```

-63-

1751 ATTATAAGCC TAGACCTGAA AAAGAAGGAA ATTTAGTTCC TAATATCTTG  
 1801 TGGGGGAATG CTGTAGATGT CAGATCCTTA ATGCAGGTTT AAGAGACCCA  
 1851 TGCATCGAGC TTACAGACAG ATCGAGGGCT GTGGATCGAT GGAATTGGGA  
 5 1901 ATTTCTFCCA TGTATCTGCC TCCGAAGACA ATATAAGGTA CCGTCATAAC  
 1951 AGCGGTGGAT ATGTTCTATC TGTAAATAAT GAGATCACAC CTAAGCACTA  
 2001 TACTTCGATG GCATTTTCCC AACTCTTTAG TAGAGACAAG GACTATGCGG  
 2051 TTTCCAACAA CGAATACAGA ATGTATTTAG GATCGTATCT CTATCAATAT  
 2101 ACAACCTCCC TAGGGAATAT TTTCCGTAT GCTTCGCGTA ACCCTAATGT  
 10 2151 AAACGTCGGG ATTCTCTCAA GAAGGTTTCT TCAAAATCCT CTTATGATTT  
 2201 TTCATTTTTT GTGTGCTTAT GGTCA TGCCA CCAATGATAT GAAAACAGAC  
 2251 TACGCAAAAT TCCCTATGGT GAAAACAGC TGGAGAAACA ATTGTTGGGC  
 2301 TATAGAGTGC GGAGGGAGCA TGCCTCTATT GGTATTTGAG AACGGAAGAC  
 2351 TTTTCCAAGG TGCCATCCCA TTTATGAAAC TACAATTAGT TTATGCTTAT  
 15 2401 CAGGGAGATT TCAAAGAGAC GACTGCAGAT GGCCGTAGAT TTAGTAATGG  
 2451 GAGTTTAAACA TCGATTTCTG TACCTCTAGG CATACGCTTT GAGAAGCTGG  
 2501 CACTTTCTCA GGATGTACTC TATGACTTTA GTTTCCTCCTA TATTCCTGAT  
 2551 ATTTTCCGTA AGGATCCCTC ATGTGAAGCT GCTCTGGTGA TTAGCGGAGA  
 2601 CTCCTGGCTT GTTCCGGCAG CACACGTATC AAGACATGCT TTTGTAGGGA  
 2651 GTGGAACGGG TCGGTATCAC TTTAACGACT ATACTGAGCT CTTATGTCGA  
 20 2701 GGAAGTATAG AATGCCGCCC CCATGCTAGG AATTATAATA TAAACTGTGG  
 2751 AAGCAAATTT CGTTTTTAG

The PSORT algorithm predicts an outer membrane location (0.921).

The protein was expressed in *E.coli* and purified both as a his-tag and GST-fusion product. The GST-fusion is shown in Figure 21A. This recombinant protein was used to immunise mice, whose sera  
 25 were used in a Western blot (Figure 21B) and for FACS analysis (Figure 21C).

This protein also showed good cross-reactivity with human sera, including sera from patients with pneumonitis.

These experiments show that cp6260 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

## 30 Example 22

The following *C.pneumoniae* protein (PID 4376456) was expressed <SEQ ID 43; cp6456>:

1 MSSPVNNTPS APNIPAPPT TPGIPTTKPR SSFIEKVIV AKYILFAIAA  
 51 TSGALGTILG LSGALTPGIG IALLVIFVSV MVLLGLILKD SISGGEERRL  
 101 REEVSRFTSE NQRLTVITTT LETEVKDLKA AKDQLTLEIE AFRNENGNLK  
 35 151 TTAEDLEEQV SKLSEQLLEAL ERINQLIQAN AGDAQEISSE LKKLISGWDS  
 201 KVVEQINTSI QALKVLGLQE WVQEAQTHVK AMQEIQIALQ AEILGMHNQS  
 251 TALQKSVENL LVQDQALTRV VGELLESENK LSQACSALRQ EIEKLAQHET  
 301 SLQQRIDAML AQEQNLAEQV TALEKMKQEA QKAESEFIAC VRDRTFGRRE  
 351 TPPPTFPVVE GDESQBEDEG GTPPVSPSS PVDRTAGDGQ \*

40 The cp6456 nucleotide sequence <SEQ ID 44> is:

1 ATGTCATCTC CTGTAAATAA CACACCCTCA GCACCAAACA TTCCAATACC  
 51 AGCGCCACAG ACTCCAGGTA TTCCTACAAC AAAACCTCGT TCTAGTTTCA  
 101 TTGAAAAGGT TATCATTGTA GCTAAGTACA TACTATTTGC AATTGCAGCC  
 45 151 ACATCAGGAG CACTCGGAAC AATTCTAGGT CTATCTGGAG CGCTAACCCC  
 201 AGGAATAGGT ATTGCCCTTC TTGTTATCTT CTTTGTCTCT ATGGTGCTTT  
 251 TAGGTTTAAT CCTTAAAGAT TCTATAAGTG GAGGAGAAGA ACGCAGGCTC  
 301 AGAAGAGAGG TCTCTCGATT TACAAGTGAG AATCAACGGT TGACAGTCAT  
 351 AACCACAACA CTTGAGACTG AAGTAAAGGA TTTAAAAGCA GCTAAAGATC  
 401 AACTTACACT TGAATTCGAA GCATTTAGAA ATGAAAACGG TAATTTAAAA  
 50 451 ACAACTGCTG AGGACTTAGA AGAGCAGGTT TCTAACTTA GCGAACAATT  
 501 AGAAGCACTA GAGCGAATTA ATCAACTTAT CCAAGCAAAC GCTGGAGATG  
 551 CTCAAGAAAT TTCGTCTGAA CTAAAGAAAT TAATAAGCGG TTGGGATTCC  
 601 AAAGTTGTTG AACAGATAAA TACTTCTATT CAAGCATTGA AAGTGTATT  
 651 GGGTCAAGAG TGGGTCAAG AGGCTCAAAC ACACGTAAAG CCAATGCAAG  
 55 701 AGCAAATTC AAGCATGCAA GCTGAAATTC TAGGAATGCA CAATCAATCT

-64-

```

751 ACAGCATTGC AAAAGTCAGT TGAGAATCTA TTAGTACAAG ATCAAGCTCT
801 AACAAGAGTA GTAGGTGAGT TGTTAGAGTC TGAGAACAAG CTAAGCCAAG
851 CTTGTTCTGC GCTACGTCAA GAAATAGAAA AGTTGGCCCA ACATGAAACA
901 TCTTTGCAAC AACGTATTGA TCGCATGCTA GCCCAAGAGC AAAATTGCGC
951 AGAGCAGGTC ACAGCCCTTG AAAAAATGAA ACAAGAAGCT CAGAAGGCTG
1001 AGTCCGAGTT CATTGCTTGT GTACGTGATC GAACTTTCGG ACGTCGTGAA
1051 ACACCTCCAC CAACAACACC TGTAGTTGAA GGTGATGAAA GTCAAGAAGA
1101 AGACGAAGGA GGTACTCCCC CAGTATCACA ACCATCTTCA CCCGTAGATA
1151 GAGCAACAGG AGATGGTCAG TAA

```

10 The PSORT algorithm predicts inner membrane (0.127).

The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 22A. The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 22B) and for FACS analysis (Figure 22C). A his-tag protein was also expressed.

15 These experiments show that cp6456 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 23

The following *C.pneumoniae* protein (PID 4376729) was expressed <SEQ ID 45; cp6729>:

```

1 MKIPLHKLLI SSTLVTPILL SIATYGADAS LSPTDSFDGA GGSFTTPKST
51 ADANGTNYVL SGNVYINDAG KGTALTGCCF TETTGDLTFT GKGYSFSFNT
20 101 VDAGSNAGAA ASTTADKALT FTGFNSLSFI AAPGTTVASG KSTLSSAGAL
151 NLTDNGTILF SQNVSNNEANN NGGAIITKTL SISGNTSSIT FTSNSAKKLG
201 GAIYSSAAAS ISGNTGQLVF MNNKETGGG ALGFEASSSI TQNSSLFFSG
251 NTATDAAGKG GAIYCEKTGE TPTLTISGNK SLTFAENSSV TQGGATCAHG
301 LDLSAAGPTL FSNNRGNTA AGKGGALALA DSGSLSLSAN QGDIITFLGNT
25 351 LTSTSAPTST RNAIYLSSA KITNLRAAQG QSIYFYDPIA SNTTGASDVL
401 TINQPSNSP LDYSGTIVFS GEKLSADEAK AADNPTSILK QPLALASGTL
451 ALKGNVELDV NGFTQTEGST LLMQPGTKLK ADTEAISLTR LVVDLSALEG
501 NKSVSJETAG ANKTITLTSP LVFQDSSGNF YESHTINQAF TQPLVVFTAA
551 TAASDIYIDA LLTSPVQTPE PHYGYQGHWE ATWADTSTAK SGTMTWVTTG
30 601 YNPNPERRAS VVPDSLWASF TDIRTLQQIM TSQANSIYQQ RGLWASGTAN
651 FFHKDKSGTN QAFRHKSYGY IVGGSIEDFS ENIFSVAFCQ LFGKDKDLFI
701 VENTSHNYLA SLYLQHRAFL GGLPMPSPFGS ITDMLKDIPL ILNAQLSYSY
751 TKNDMDTRYT SYPEAQGSWT NNSGALELGG SLALYLPKRA PFFQGYFPFL
801 KFAQVYSRQQ NFKESGAEAR AFDDGDLVNC SIPVGIRLEK ISEDEKNNFE
35 851 ISLAYIGDVY RKNPRSRTSL MVSGASWTSI CKNLARQAFI ASAGSHLTLS
901 PHVELSGEAA YELRGSATY NVDCGLRYSF *

```

A predicted signal peptide is highlighted.

The cp6729 nucleotide sequence <SEQ ID 46> is:

```

1 ATGAAAATAC CCTTGCACAA ACTCCTGATC TCTTCGACTC TTGTCACTCC
40 51 CATTCCTATTG AGCATTGCAA CTTACGGAGC AGATGCTTCT TTATCCCTTA
101 CAGATAGCTT TGATGGAGCG GCGGCTCTA CATTTACTCC AAAATCTACA
151 GCAGATGCCA ATGGAACGAA CTATGTCTTA TCAGGAAATG TCTATATAAA
201 CGATGCTGGG AAAGGCACAG CATTAACAGG CTGCTGCTTT ACAGAACTA
251 CGGGTGATCT GACATTACTT GGAAAGGGAT ACTCATTTTC ATTCAACACG
45 301 GTAGATGCGG GTTCGAATGC AGGAGCTGCG GCAAGCACAA CTGCTGATAA
351 AGCCCTAACA TTCACAGGAT TTTCTAACCT TTCTTCAATT GCAGCTCCTG
401 GAACTACAGT TGCTTCAGGA AAAAGTACTT TAAGTTCTGC AGGAGCCTTA
451 AATCTTACCG ATAATGGAAC GATTCTCTTT AGCCAAAACG TCTCCAATGA
501 AGCTAATAAC AATGGCGGAG CGATCACCAC AAAAATCTTT TCTATTCTTG
50 551 GGAATACCTC TTCTATAACC TTCCTAGTA ATAGCGCAA AAAATTAGGT
601 GGAGCGATCT ATAGCTCTGC GGCTGCAAGT ATTTAGGAA ACACCGGCCA
651 GTTAGTCTTT ATGAATAATA AAGGAGAAAC TGGGGGTGGG GCTCTGGGCT
701 TTGAAGCCAG CTCCTCGATT ACTCAAAATA GCTCCCTTTT CTTCTCTGGA
751 AACACTGCAA CAGATGCTGC AGGCAAGGGC GGGGCCATTT ATTGTGAAAA
55 801 AACAGGAGAG ACTCTACTC TTAATATCTC TGGAAATAAA AGTCTGACCT
851 TCGCCGAGAA CTCCTCAGTA ACTCAAGGCG GAGCAATCTG TGCCCATGGT

```

5 901 CTAGATCTTT CCGCTGCTGG CCCTACCCCTA TTTTCAAATA ATAGATGCGG  
 951 GAACACAGCT GCAGGCAAGG GCGGCGCTAT TGCAATTGCC GACTCTGGAT  
 1001 CTTTAAAGTCT CTCTGCAAAT CAAGGAGACA TCACGTTCCCT TGGCAACACT  
 1051 CTAACCTCAA CCTCCGCGCC AACATCGACA CGGAATGCTA TCTACCTGGG  
 1101 ATCGTCAGCA AAAATTACGA ACTTAAGGGC AGCCCAAGGC CAATCTATCT  
 1151 ATTTCTATGA TCCGATTGCA TCTAACACCA CAGGAGCTTC AGACGTTCTG  
 1201 ACCATCAACC AACCGGATAG CAACTCGCCT TTAGATTATT CAGGAACGAT  
 1251 TGTATTTTCT GGGGAAAAGC TCTCTGCAGA TGAAGCGAAA GCTGCTGATA  
 1301 ACTTCACATC TATATTAAG CAACCATTGG CTCTAGCCTC TGGAACCTTA  
 1351 GCACTCAAAG GAAATGTCGA GTTAGATGTC AATGGTTTCA CACAGACTGA  
 1401 AGGCTCTACA CTCCTCATGC AACCAGGAAC AAAGCTCAAA GCAGATACTG  
 1451 AAGCTATCAG TCTTACCAAA CTTGTCTGTG ATCTTTCTGC CTTAGAGGGA  
 1501 AATAAGAGTG TGTCCATTGA AACAGCAGGA GCCAACAAAA CTATAACTCT  
 1551 AACCTCTCCT CTTGTTTTCC AAGATAGTAG CGGCAATTTT TATGAAAGCC  
 1601 ATACGATAAA CCAAGCCTTC ACGCAGCCTT TGGTGGTATT CACTGCTGCT  
 1651 ACTGCTGCTA GCGATATTTA TATCGATGCG CTTCTCACIT CTCCAGTACA  
 1701 AACTCCAGAA CCTCATTACG GGTATCAGGG ACATTGGGAA GCCACTTGGG  
 1751 CAGACACATC AACTGCAAAA TCAGGAATA TGAATTGGGT AACTACGGGC  
 1801 TACAACCCCTA ATCTTGAGCG TAGAGCTTCC GTAGTTCCCG ATTCAATTATG  
 20 1851 GGCATCCTTT ACTGACATTC GCACTCTACA GCAGATCATG ACATCTCAAG  
 1901 CGAATAGTAT CTATCAGCAA CGAGGACTCT GGGCATCAGG AACTGCGAAT  
 1951 TTCTTCCATA AGGATAAATC AGGAACTAAC CAAGCATTCG GACATAAAAG  
 2001 CTACGGCTAT ATTGTGGAG GAAGTGCTGA AGATTTTCTT GAAAATATCT  
 2051 TCAGTGTAGC TTTCTGCCAG CTCCTCGGTA AAGATAAAGA CCTGTTTATA  
 25 2101 GTTGAAAATA CCTCTCATAA CTATTTAGCG TCGCTATACC TGCAACATCG  
 2151 AGCATTCCTA GGAGGACTTC CCATGCCCTC ATTTGGAAAT ATCACCAGCA  
 2201 TGCTGAAAGA TATTCTCTC ATTTTGAATG CCCAGCTAAG CTACAGCTAC  
 2251 ACTAAAAATG ATATGGATAC TCGCTATACT TCCTATCCTG AAGCTCAAGG  
 2301 CTCTTGGACC AATAACTCTG GGGCTCTAGA GCTCGGAGGA TCTCTGGCTC  
 30 2351 TATATCTCCC TAAAGAAGCA CCGTCTCTCC AGGGATATTT CCCCTTCTTA  
 2401 AAGTTCAGG CAGTCTACAG CCGCCAACAA AACTTTAAAG AGAGTGGCGC  
 2451 TGAAGCCCGT GCTTTTGATG ATGGAGACCT AGTGAAGTGC TCTATCCCTG  
 2501 TCGGCATTCT GTTAGAAAA ATCTCCGAAG ATGAAAAAAT TAATTTCGAG  
 2551 ATTTCTCTAG CCTACATTGG TGATGTGTAT CGTAAAAATC CCCGTTCCGG  
 35 2601 TACTTCTCTA ATGGTCAATG GAGCCTCTTG GACTTCGCTA TGTAAAAACC  
 2651 TCGCACGACA AGCCTTCTTA GCAAGTGCTG GAAGCCATCT GACTCTCTCC  
 2701 CCTCATGTAG AACTCTCTGG GGAAGCTGCT TATGAGCTTC GTGGCTCAGC  
 2751 ACACATCTAC AATGTAGATT GTGGGCTAAG ATACTCATTC TAG

The PSORT algorithm predicts outer membrane (0.927).

- 40 The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 23A. The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 23B) and for FACS analysis (Figure 23C). A his-tag protein was also expressed.

The cp6729 protein was also identified in the 2D-PAGE experiment (Cpn0446) and showed good cross-reactivity with human sera, including sera from patients with pneumonitis.

- 45 These experiments show that cp6729 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

#### Example 24

The following *C.pneumoniae* protein (PID 4376849) was expressed <SEQ ID 47; cp6849>:

50 1 MSKLIRRVVT VLALTSMA SC FASGGIEAAV AESLITKIVA SAETKPAPVP  
 51 MTAKKVRLLR RNKQPVBEQS RGAFCDKEFY PCEEGRCQPV EAQQESCYGR  
 101 LYSVKVNDDC NVEICQSVPE YATVGSYPPI EILAIGKKDC VDVVITQQLP  
 151 CEAEFVSSDP ETTPTSDGKL VWKIDRLGAG DKCKITVWVK PLKEGCCFTA  
 201 ATVCACPELR SYTKCQPAI CIKQEGPDCA CLRCPCYKI EVVNTGSAIA  
 251 RNVTVDNFVP DGYSHASGQR VLSFNLGDMR PGDKKVFTVE FCPQRRGQIT  
 55 301 NVATVITYCG HKCSANVTTV VNEPCVQVNI SGADWSYVCK PVEYSISVSN  
 351 PGDLVLHDVV IQDTLPSGVT VLEAPGGEIC CNKVWRIKE MCPGETLQFK



-66-

401 LVVKAQVPGR FTNQVAVTSE SNCGTCTSCA ETTTHWKGLA ATHMCVLDTN  
 451 DPICVGENTV YRICVTNRGS AEDTNVSLIL KFSKELQPIA SSGPTKGTIS  
 501 GNTVVPDALP KLGSKESVERF SVTLKGIAPG DARGEAILSS DTLTSPVSDT  
 551 ENTHVY\*

5 A predicted signal peptide is highlighted.

The cp6849 nucleotide sequence <SEQ ID 48> is:

1 ATGTCCAAAC TCATCAGACG AGTAGTTACG GTCCTTGCGC TAACGAGTAT  
 51 GGCGAGTTGC TTTGCCAGCG GGGGTATAGA GGCCGCTGTA GCAGAGTCTC  
 101 TGATTACTAA GATCGTCGCT AGTGCGGAAA CAAAGCCAGC ACCTGTTCCCT  
 151 ATGACAGCGA AGAAGGTTAG ACTTGTCCGT AGAAATAAAC AACCAGTTGA  
 201 ACAA AAAAGC CGTGGTGTCT TTTGTGATAA AGAATTTTAT CCCTGTGAAG  
 251 AGGGACGATG TCAACCTGTA GAGGCTCAGC AAGAGTCTTG CTACGGAAGA  
 301 TTGTATTCTG TAAAAGTAAA CGATGATGTC AACGTAGAAA TTTGCCAGTC  
 351 CGTTCCAGAA TACGCTACTG TAGGATCTCC TTACCCCTATT GAAATCCTTG  
 15 401 CTATAGGCAA AAAAGATTGT GTTGATGTTG TGATTACACA ACAGCTACCT  
 451 TCGGAAGCTG AATTCGTAAG CAGTGATCCA GAAACAACCT CTACAAGTGA  
 501 TGGGAAATTA GTCTGAAAAA TCGATCGCCT GGGTGCAGGA GATAAATGCA  
 551 AAATTACTGT ATGGGTAAAA CCTCTTAAAG AAGGTTGCTG CTTACAGCT  
 601 GCTACTGTAT GTGCTTGCCC AGAGCTCCGT TCTTATACTA AATGCGGTCA  
 20 651 ACCAGCCATT TGTAATTAAGC AAGAAGGACC TGACTGTGCT TGCCTAAGAT  
 701 GCCCTGTATG CTACAAAATC GAAGTAGTGA ACACAGGATC TGCTATTGCC  
 751 CGTAACGTAA CTGTAGATAA TCCTGTTCCC GATGGCTATT CTCATGCATC  
 801 TGGTCAAAGA GTTCTCTCTT TTAACCTTAGG AGACATGAGA CCTGGCGATA  
 851 AAAAGGTATT TACAGTTGAG TTCTGCCCTC AAAGAAGAGG TCAAACTCACT  
 25 901 AACGTTGCTA CTGTAACCTA CTGCGGTGGA CACAAATGTT CTGCAAATGT  
 951 AACTACAGTT GTTAATGAGC CTTGTGTACA AGTAAATATC TCTGGTGCTG  
 1001 ATTTGGTCTTA CGTATGTAAA CCTGTGGAGT ACTCTATCTC AGTATCGAAT  
 1051 CCTGGAGACT TGGTTCTTCA TGATGTCGTG ATCCAAGATA CACTCCCTTC  
 1101 TGGTGTTACA GTACTCGAAG CTCCTGGTGG AGAGATCTGC TGTAAATAAG  
 30 1151 TTGTTTGGCG TATTAAAGAA ATGTGCCCAG GAGAAACCTT CCAGTTTAAA  
 1201 CTTGTAGTGA AAGCTCAAGT TCCTGGAAGA TTCACAAATC AAGTTGCAGT  
 1251 AACTAGTGAG TCTAACTGCG GAACATGTAC ATCTTGCCGA GAAACAACAA  
 1301 CACATTGGAA AGGCTTTCGA GCTACCCATA TGTGCGTATT AGACACAAAT  
 1351 GATCCTATCT GTGTAGGAGA AAATACTGTC TATCGTATCT GTGTAATAA  
 35 1401 CCGTGGTTCT GCTGAAGATA CTAACGTATC TTTAATCTTG AAGTTCTCAA  
 1451 AAGAACTTCA GCCAATAGCT TCTTCAGGTC CAACTAAAGG AACGATTTC  
 1501 GGTAAATACG TTGTTTTCGA CGCTTTACCT AAACTCGGTT CTAAGGAATC  
 1551 TGTAGAGTTT TCTGTTACCT TGAAAGGTAT TGCTCCCGGA GATGCTCGCG  
 1601 GCGAAGCTAT TCTTTCTTCT GATACACTGA CTTACCAGT ATCAGACACA  
 40 1651 GAAATATACC ACGTGTATTA A

The PSORT algorithm predicts periplasmic space (0.93).

The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 24A, and also as a his-tag protein. The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (Figure 24B) and for FACS analysis (Figure 24C).

45 The cp6849 protein was also identified in the 2D-PAGE experiment (Cpn0557).

These experiments show that cp6849 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 25

The following *C.pneumoniae* protein (PID 4376273) was expressed <SEQ ID 49; cp6273>:

50 1 MGLFHLTLFG LLLCSLPISL VAKFPESVGH KILYISTQST QQALATYLEA  
 51 LDAYGDHDFV VLRRKIGEDYL KQSIHSSDPQ TRKSTIIGAG LAGSSEALDV  
 101 LSQAMETADP LQQLLVLSAV SGHLGKTSDD LLFKALASPY PVIRLEAAYR  
 151 LANLKNTKVI DHLHSFIHKL PEEIQCLSAA IFLRLETEES DAYIRDLLAA  
 201 KKSARSATA LQIGBYQQR FLPTLRNLLT SASPDQAEI LYALGKLKDG

-67-

251 QSYYNIKKQL QKPDVDVTLA AAQALIALGK EEDALPVIKK QALEERPRAL  
 301 YALRHLPSSEI GIPIALPIFL KTKNSEAKLN VALALLEIGC DTPKLLLEYIT  
 351 ERLVQPHYNE TLALSFSGKR TLQNWKRVNI IVPQDPQERE RLLSTTRGLE  
 401 EQILTFLEFRL PKEAYLPCYI KLLASQKTQL ATTAISFLSH TSHQEALDLL  
 451 FQAAKLPGEP IIRAYADLAI YNLTKDPEKK RSLHDYAKKL IQETLLFVDT  
 501 ENQRPHPSMP YLRYQVTPES RTKLMLDILE TLATSKSSED IRLLIQLMTE  
 551 GDAKNFPVLA GLLIKIVE\*

A predicted signal peptide is highlighted.

The cp6273 nucleotide sequence <SEQ ID 50> is:

10 1 ATGGGACTAT TCCATCTAAC TCTCTTTGGA CTTTATTTGT GTAGTCTTCC  
 51 CATTTCCTCTT GTTGCTAAAT TCCCTGAGTC TGTAGGTCAT AAGATCCTTT  
 101 ATATAAGTAC GCAATCTACA CAGCAGGCC TTAGCAACATA TCTGGAAGCT  
 151 CTAGATGCCT ACGGTGATCA TGACTTCTTC GTTTTAAAGAA AAATCGGAGA  
 201 AGACTATCTC AAGCAAAGCA TCCACTCCTC AGATCCGCAA ACTAGAAAAA  
 15 251 GCACCATCAT TGGAGCAGGC CTGGCGGGAT CTTCAGAAGC CTTGGACGTG  
 301 CTCTCCCAAG CTATGGAAAC TGCAGACCCC CTGCAGCAGC TACTGGTTTT  
 351 ATCGGCAGTC TCAGGACATC TTGGGAAAAC TTCTGACGAC TTACTGTTTA  
 401 AAGCTTTAGC ATCTCCCTAT CCTGTCTATC GCTTAGAAGC CGCCTATAGA  
 451 CTTGCTAATT TGAAGAACAC TAAAGTCATT GATCATCTAC ATTCTTTCAT  
 20 501 TCATAAGCTT CCCGAAGAAA TCCAATGCCT ATCTGCGGCA ATATTCCTAC  
 551 GCTTGGAGAC TGAAGAATCT GATGCTTATA TTCGGGATCT CTTAGCTGCC  
 601 AAGAAAAGCG CGATTTCGGAG TGCCACAGCT TTGCAGATCG GAGAATACCA  
 651 ACAAAAACGC TTTCTTCCGA CACTTAGGAA TTTGCTAACG AGTGCCTCTC  
 701 CTCAAGATCA AGAAGCTATT CTTTATGCTT TAGGGAAGCT TAAGGATGGT  
 25 751 CAGAGCTACT ACAATATAAA AAAGCAATTG CAGAAGCCTG ATGTGGATGT  
 801 CACTTTAGCA GCAGCTCAAG CTTTAAATTGC TTTGGGGAAA GAAGAGGACG  
 851 CTCTTCCCGT GATAAAAAAG CAAGCACTTG AGGAGCGGCC TCGAGCCCTG  
 901 TATGCCCTTAC GGCATCTACC CTCTGAGATA GGGATTCCGA TTGCCCTGCC  
 951 GATATTCCTA AAAACFAAGA ACAGCGAAGC CAAGTTGAAT GTAGCTTTAG  
 30 1001 CTCTCTTAGA GTTAGGGTGT GACACCCCTA AACTACTGGA ATACATTACC  
 1051 GAAAGGCTTG TCCAACCACA TTATAATGAG ACTCTAGCCT TGAGTTTCTC  
 1101 TAAGGGCGCT ACTTTACAAA ATTGGAAGCG GGTGAACATC ATAGTCCCTC  
 1151 AAGATCCCA GAGAGGGGAA AGGTGCTCT CCACAACCCG AGGTCTTGAA  
 1201 GAGCAGATCC TTACGTTTCT CTTCCGCTTA CCTAAAGAAG CTTACCTCCC  
 35 1251 CTGTATTTAT AAGCTTTTGG CGAGTCAGAA AACTCAGCTT GCCACTACTG  
 1301 CGATTTCTTT TTTAAGTCAC ACCTCACATC AGGAAGCCTT AGATCTACTT  
 1351 TTCCAAGCTG CGAAGCTTCC TGGAGAACCT ATCATCCGCG CCTATGCAGA  
 1401 TCTTGCTATT TATAATCTCA CCAAAGATCC TGAATAAAAA CGTTCTCTCC  
 1451 ATGATTATGC AAAAAAGCTA ATTCAGGAAA CCTTGTATT TGTGGACACG  
 40 1501 GAAAACCAAA GACCCCATCC CAGCATGCC TATCTACGTT ATCAGGTCAC  
 1551 CCCAGAAAAGC CGTACGAAGC TCATGTTGGA TATTCTAGAG ACACTAGCCA  
 1601 CCTCGAAGTC TTCCGAAGAT ATCCGTTTAT TGATACAAC GATGACGGAA  
 1651 GGAGATGCAA AAAATTTCCT AGTCCCTGCA GGCTTACTCA TAAAAATTGT  
 1701 GGAGTAA

45 The PSORT algorithm predicts a periplasmic location (0.922).

The protein was expressed in *E.coli* and purified as a his-tag product and as a GST-fusion product, as shown in Figure 25A. The recombinant GST-fusion was used to immunise mice, whose sera were used in a Western blot (Figure 25B) and for FACS analysis (Figure 25C).

50 This protein also showed good cross-reactivity with human sera, including sera from patients with pneumonitis.

These experiments show that cp6273 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

#### Example 26

The following *C.pneumoniae* protein (PID 4376735) was expressed <SEQ ID 51; cp6735>:

-68-

1 MTILRNFLTC SALFLALPAA AQVVYLHESD GYNGAINNKS LEPKITCYPE  
 51 GTSYIFLDDV RISNVKHDQE DAGVFINRSG NLFFMGNRCN FTFHNLMTGEG  
 101 FGAATISNRVG DTTLTLSNFS YLAFTSAPLL PQGGGAIYSL GSVMIENSEE  
 151 VTFCGNYSSW SGAAIYTPYL LGSKASRPSV NLSGNRYLVF RDNVSQGYGG  
 5 201 AISTHNLTLT TRGPSCFENN HAYHDVNSNG GAIAIAPGGS ISISVKSGDL  
 251 IFKGNITASQD GNTIHNSIHL QSGAQFKNLR AVSESGVYFY DPISHSESHK  
 301 ITDLVINAPE GKETYEGTIS FSGLCDDHE VCAENLTSTI LQDVTLAGGT  
 351 LSLSDGVTLQ LHSFKQEQASS TLTMSPGTTL LCSGDARVQN LHILIEDTDN  
 401 FVPVRIRAED KDALVSLEKL KVAFEAYWSV YDFPQFKEAF TIPILLELLGP  
 10 451 SFDSLLLGET TLERTQVTTE NDAVRGFWSL SWEYPPSLD KDRRITPTTK  
 501 TVFLTWNPEI TSTP\*

A predicted signal peptide is highlighted.

The cp6735 nucleotide sequence <SEQ ID 52> is:

1 ATGACCATAC TTCGAAATTT TCTTACCTGC TCGGCTTTAT TCCTCGCTCT  
 15 51 CCCTGCAGCA GCACAAGTTG TATATCTTCA TGAAAGTGAT GGTATAACG  
 101 GTGCTATCAA TAATAAAAGC TTAGAACCTA AAATTACCTG TTATCCAGAA  
 151 GGAAGTCTTT ACATCTTCTT AGATGACGTG AGGATTCCA ACGTTAAGCA  
 201 TGATCAAGAA GATGCTGGGG TTTTATAAAA TCGATCTGGG AATCTTTTTT  
 251 TCATGGGCAA CCGTTGCAAC TTCACTTTTC ACAACCTTAT GACCGAGGGT  
 20 301 TTTGGCGCTG CCAITTCGAA CCGCGTTGGA GACACCACTC TCACTCTCTC  
 351 TAATTTTTCT TACTTAGCGT TCACCTCAGC ACCTCTACTA CCTCAAGGAC  
 401 AAGGAGCGAT TTATAGTCTT GGTTCCTGTA TGATCGAAAA TAGTGAGGAA  
 451 GTGACTTTCT GTGGGAACATA CTCTTCGTGG AGTGGAGCTG CGATTATATC  
 501 TCCCTACCTT TTAGGTCTTA AGGCGAGTCG TCCTTCAGTA AATCTCAGCG  
 25 551 GGAACCGCTA CCTGGTGTTT AGAGACAATG TGAGCCAAGG TTATGGCGGC  
 601 GCCATATCTA CCCACAATCT CACACTCACG ACTCGAGGAC CTTCTGTGTTT  
 651 TGAAAAATAAT CATGCTTATC ATGACGTGAA TAGTAATGGA GGAGCCATTG  
 701 CCAATGCTCC TGGAGGATCG ATCTCTATAT CCGTGAAAAG CGGAGATCTC  
 751 ATCTTCAAAG GAAATACAGC ATCACAAGAC GGAAATACAA TACACAATC  
 30 801 CATCCATCTG CAATCTGGAG CACAGTTTAA GAACCTACGT GCTGTTTCAG  
 851 AATCCGGAGT TTAITTTCTAT GATCCTATAA GCCATAGCGA GTCGCATAAA  
 901 ATTACAGATC TTGTAATCAA TGCTCCTGAA GGAAAGGAAA CTTATGAAGG  
 951 AACAAATTAGC TTCTCAGGAC TATGCCTGGA TGATCATGAA GTTTGTGCGG  
 1001 AAAATCTTAC TTCCACAATC CTACAAGATG TCACATTAGC AGGAGGAACT  
 35 1051 CTCTCTCTAT CGGATGGGGT TACCTTGCAA CTGCATTCTT TTAAGCAGGA  
 1101 AGCAAGCTCT ACGCTTACTA TGTCTCCAGG AACCACCTCTG CTCTGCTCAG  
 1151 GAGATGCTCG GGTTCAGAAAT CTGCACATCC TGATTGAAGA TACCGACAAC  
 1201 TTTGTTCTCG TAAGGATTCTG CGCCGAGGAC AAGGATGCTC TTGTCTCATT  
 1251 AGAAAAACTT AAAGTTGCCT TTGAGGCTTA TTGGTCCGTC TATGACTTTC  
 40 1301 CTCAATTTAA GGAAGCCTTT ACGATTCTCT TTCTTGAAC TCTAGGCCTT  
 1351 TCTTTTGACA GTCTTCTCCT AGGGGAGACC ACTTTGGAGA GAACCAAGT  
 1401 CACAACAGAG AATGACGCCG TTCGAGGTTT CTGGTCCCTA AGCTGGGAAG  
 1451 AGTACCCCCC TTCCTCTGGAT AAAGACAGAA GGATCACACC AACTAAGAAA  
 1501 ACTGTTTTCC TCACTTGGA TCTTGAGATC ACTTCTACGC CATAA

45 The PSORT algorithm predicts an outer membrane location (0.922).

The protein was expressed in *E.coli* and purified as a his-tag product and as a GST-fusion product, as shown in Figure 26A. The recombinant GST-fusion protein was used to immunise mice, whose sera were used in a Western blot (Figure 26B).

50 These experiments show that cp6735 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 27

The following *C.pneumoniae* protein (PID 4376784) was expressed <SEQ ID 53; cp6784>:

1 MNRRKARVVV ALFAMTALIS VGCCPWSQAK SRCSDIKYIP VVNRLLEVCV  
 55 51 LPEAENVEDL IESSAWVLT PEERFSGELV SICQVKDEHA FYNDLSLLHM  
 101 TQAVPSYSAT YDCAVVFGGP LPALRQRLDF LVREWQGVV FKKIVFLCGE  
 151 RGRYQSIEEQ EHFDSRYNP FPTEENWESG NRVTPSSEEE IAKFVWMQML

-69-

201 LPRAWRDSTS GVRVTFLLAK PEENRVVANR KDTLLLFRRSY QEAFPGRVLF  
 251 VSSQPFIGLD ACRVGQFFKG ESYDLAGPGF AQGVLYKHYWA PRICLHTLAE  
 301 WLKETNGCLN ISEGCPG\*

A predicted signal peptide is highlighted.

5 The cp6784 nucleotide sequence <SEQ ID 54> is:

1 ATGAATAGAA GAAAAGCAAG ATGGGTAGTG GCATTGTTTCG CAATGACGGC  
 51 GCTCATTTCT GTTGGGTGTT GTCCTTGGTC ACAAGCGAAA TCAAGATGTT  
 101 CTATTGATAA GTATATTCCT GTAGTCAATC GTTTACTAGA AGTTTGTTGA  
 151 CTTCCTGAAG CTGAGAATGT TGAGGATTTA ATCGAGTCCT CGTCGTCTTG  
 201 GGTACTGACT CCTGAAGAAC GTTTTCTTGG AGAGTTAGTC TCTATCTGTC  
 251 AGGTTAAAGA TGAGCATGCT TTCTATAACG ATTTGTCTTT ATTACATATG  
 301 ACTCAGGCTG TGCCTTCGTA TTCTGCAACG TATGATTGTC CTGTAGTTTT  
 351 TGGCGGGCCT TTGCCAGCGC TACGTCAGCG CTTAGATTTT TTGGTGCGAG  
 401 AGTGGCAGCG TGGCGTGCGC TTTAAGAAAA TCGTTTTCCT ATGTGGAGAG  
 451 CGAGGGCGCT ATCAGTCTAT TGAAGAACAA GAGCATTTCT TTGATCTCTG  
 501 GTACAATCCT TTCCCTACTG AAGAGAAGCTG GGAATCTGGT AACCGAGTTA  
 551 CTCCCTCTTC TGAAGAAGAG ATTGCCAAAT TTGTTTGGAT GCAATGCTTT  
 601 TTACCTAGAG CATGGCGAGA TAGTACTTCA GGAGTCAGAG TGACATTTCT  
 651 TCTAGCAAAG CCAGAGGAAA ATCGTGTGGT TGCGAATCGT AAGGACACCT  
 701 TACTTTTATT CCGTCTCTAT CAAGAAGCGT TTCCGGGACG CGTGTATTTT  
 751 GTAAGTAGTC AACCTTTTAT CCGTTTAGAT GCTTGCAGGG TCGGGCAGTT  
 801 TTTCAAAGGG GAAAGCTATG ATCTTGCTGG ACCTGGATTT GCTCAAGGAG  
 851 TCTTGAAGTA TCATTGGGCT CCAAGGATTT GTCTACATAC TTTAGCGGAA  
 901 TGGTTAAAGG AAACGAACGG CTGCTTAAAT ATTTTCAGAGG GTTGTTTTGG  
 951 ATGA

The PSORT algorithm predicts a periplasmic location (0.894).

The protein was expressed in *E. coli* and purified as a his-tag product and as a GST-fusion product, as shown in Figure 27A. The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (Figure 27B). The GST-fusion product was used for FACS analysis (Figure 27C).

30 The cp6784 protein was also identified in the 2D-PAGE experiment (Cpn0498).

These experiments show that cp6784 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 28

The following *C. pneumoniae* protein (PID 4376960) was expressed <SEQ ID 55; cp6960>:

35 1 MNRRWNVLVA TVALALSVAS CDVRSKDKDK DQGSLEYKD NKDTNDIELS  
 51 DNQKLSRTFG HLLARQLRKS EDMFFDIAEV AKGLQAEIVC KSAPLTETET  
 101 EEKMAEVQKL VFEKRSKENL SLAEKFLKEN SKNAGVVEVQ PSKLQYKIIK  
 151 EGAGKAISGR PSALLHYKGS FINGQVFSSS EGNNEPILLP LGQTIPGFAL  
 201 GMQGMKEGET RVLYIHPDLA YGTAGQLPPN SLLIFBINLI QASADEVA  
 40 251 PQEGNQGE\*

A predicted signal peptide is highlighted.

The cp6960 nucleotide sequence <SEQ ID 56> is:

1 ATGAACAGAC GGTGGAATTT AGTTTGTAGCA ACAGTAGCTC TGGCACTCTC  
 45 51 CGTCGCTTCT TGTGACGTAC GGTCTAAGGA TAAAGACAAG GATCAGGGGT  
 101 CGTTAGTGGG ATATAAGAT AACAAAGATA CCAATGACAT AGAATTATCC  
 151 GATAATCAAA AGTTATCCAG AACATTGGT CATTATATAG CACGCCAATT  
 201 ACGCAAGTCA GAAGATATGT TTTTGTATAT TGCAGAAGTG GCTAAGGGGT  
 251 TGCAGCGCGA ATTGGTTTGT AAAAGTGCTC CTTTAACAGA AACAGAGTAT  
 301 GAAGAAAAAA TGGCTGAAGT ACAGAAGTTG GTTTTGTAAA AAAAATCAAA  
 351 AGAAAAATCTT TCATTGGCAG AAAAATTTCT AAAAGAAAAAT AGCAAGAACG  
 50 401 CTGGTGTGTG TGAAGTGCAA CCAAGTAAAT TGCAATACAA AATTATTA

-70-

451 GAAGGTGCAG GGAAAGCAAT TTCAGGTAAA CCTTCAGCTC TATTGCACTA  
 501 CAAGGGTTC TCCATCAATG GCCAAGTATT TAGCAGTTCA GAAGGCAACA  
 551 ATGAGCCTAT CTTCGTTCTCT CTAGGCCAAA CAATTCTCTGG TTTTGCCTTA  
 601 GGTATGCAGG GCATGAAAGA AGGAGAAACT CGAGTTCTCT ACATCCATCC  
 651 TGATCTTGCT TACGGAACCG CAGGACAACT TCCTCCAAAC TCTTTATTAA  
 701 TTTTGTAAAT TAACTTGATT CAGGCTTCAG CAGATGAAGT TGCTGCTGTA  
 751 CCCCAAGAAG GAAATCAAGG TGAATGA

The PSORT algorithm predicts periplasmic space location (0.930).

The protein was expressed in *E.coli* and purified as a his-tag product and as a GST-fusion product, as shown in Figure 28A. The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (Figure 28B) and for FACS analysis (Figure 28C).

The cp6960 protein was also identified in the 2D-PAGE experiment.

These experiments show that cp6960 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 29

The following *C.pneumoniae* protein (PID 4376968) was expressed <SEQ ID 57; cp6968>:

1 MKFLLYVPLL LVLVSTGCD A KPVSFEPFSG KLSTQRFEPQ HSAREYFSQG  
 51 QEFLKKG NFR KALLCPGIIT HHFPRDILRN QAQYLIGVCY FTQDHPDLAD  
 101 KAFASYLQLP DAEYSEELFQ MKYAIQRF A QGKRKRICRL EGFPKLMNAD  
 151 EDALRIYDEI LTAFP SKDLG AQALYSKAAL LIVKNDL TEA TKTLKKLTLQ  
 201 FPLHILSSEA FVRLSEIYLQ QAKKEPHNLQ YLHFAKLNEE AMKKQHPNHP  
 251 LNEVVSANVG AMREHYARGL YATGRFY EKK KKAEEAANIYY RTAITNYPDT  
 301 LLVAKCQKRL DRISKHTS\*

A predicted signal peptide is highlighted.

The cp6968 nucleotide sequence <SEQ ID 58> is:

1 ATGAAATTTC TATTATACGT TCCACTTCTT CTGTGTTCTCG TATCTACGGG  
 51 GTGCGATGCA AAACCTGTTT CTTTGTAGCC CTTTTCAGGA AAGCTTTCCA  
 101 CCCAGCGTTT TGAGCCTCAG CACTCTGCTG AAGAATATTT TTCTCAGGGA  
 151 CAGGAATTCT TAAAAAAGG AATTTTCAGA AAAGCTTTTAC TATGCTTTGG  
 201 AATCATTTACG CATCACATTCC CTAGGGACAT CTGCGTAAT CAAGCACAGT  
 251 ATCTTTATAGG AGTCTGTTAC TTCACGCAGG ATCACCAGGA TTTAGCAGAC  
 301 AAGGCATTTC CATCTTACTT ACAACTTCCT GATGCGGAGT ACTCTGAAGA  
 351 GTTGTTCAG ATGAAATATG CGATTGCTCA AAGATTGCTT CAAGGGAAGC  
 401 GTAAACGGAT TTGTCGATTA GAGGGCTTCC CAAAACTAAT GAATGCTGAT  
 451 GAAGATGCGC TACGCATTTA TGACGAGATT CTAACAGCGT TTCCTAGTAA  
 501 AGACTTAGGA GCTCAGGCC TCTATAGTAA AGCTGCGTTA CTTATTGTAA  
 551 AAAACGATCT TACAGAAAGC ACCAAAACCT TAAAAAACT CACGTTACAA  
 601 TTTCTCTTAC ATATTTTATC TTCAGAGGCC TTTGTACGTT TATCGGAAAT  
 651 CTATTTACAG CAAGCTAAGA AAGAGCCTCA CAATCTTCAA TATCTTCATT  
 701 TTGCAAAGCT TAATGAAGAG GCAATGAAAA AGCAGCATCC TAACCATCCT  
 751 CTGAATGAGG TTGTTTCTGC TAATGTTGGA GCTATGCGGG AACATTATGC  
 801 TCGAGGTTTG TATGCCACAG GTCGTTTCTA TGAGAAGAAG AAAAAAGCCG  
 851 AGGCTGCGAA TATCTATTAC CGCACTGCGA TTACAACTA CCCAGACACT  
 901 TTATTAGTGG CTAAATGTCA AAAGCGTCTA GATAGAATAT CTAAGCATAC  
 951 TTCCTAA

The PSORT algorithm predicts an inner membrane location (0.790).

The protein was expressed in *E.coli* and purified as a his-tag product and as a GST-fusion product, as shown in Figure 29A. The recombinant GST-fusion was used to immunise mice, whose sera were used in a Western blot (Figure 29B) and for FACS analysis (Figure 29C).

This protein also showed good cross-reactivity with human sera, including sera from patients with pneumonitis.

These experiments show that cp6968 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### 5 Example 30

The following *C.pneumoniae* protein (PID 4376998) was expressed <SEQ ID 59; cp6998>:

```

1  MKKLLKSALL SAAPAGSVGS LQALPVGNPS DPSLLIDGTI WEGAAGDPCD
51  PCATWCDATS LRAGFYGDYV FDRILKVDAP KTF5MGAKPT GSAAANYTTA
101 VDRPNPAYNK HLHDAEWFTN AGFIALNIWD RFDVFCFLGA SNGYIRGNST
151 AFNLVGLFGV KGTTVNANEL PNVSL5NGVV ELYTDT5FSW SVGARGALWE
201 CGCATLGAEF QYAQSKPKVE ELNVICNV5Q FSVNKP5GYK GVA5PLPTDA
251 GVATATG5TKS ATINYHEWQV GASLSYRLNS LVPYIGVQWS RATFDADNIR
301 IAQPKLPTAV LNLTAWNPSL LGNATALSTT DSFSDFMQIV SCQINKFKSR
351 KACGVTVGAT LVDADKWSLT AEARLINERA AHVSGQFRF*
```

15 A predicted signal peptide is highlighted.

The cp6998 nucleotide sequence <SEQ ID 60> is:

```

1  ATGAAAAAAC TCTTAAAGTC GGC5GTTATTA TCCGCCGCAT TTGCTGGTTC
51  TGT5TGGCTCC TTACAAGCCT TGCC5TGTAGG GAACCC5TCT GATCCAAGCT
101 TATTAATTGA TGGTACAATA TGGGAAGGTG CTGCAGGAGA TCCT5TGC5AT
151 CCTTGC5GTA CTTGGT5GCA CGCTAT5TAGC TTACGT5GCTG GATTT5ACGG
201 AGACTATGTT TTCGAC5CGTA TCTTAA5AGT AGATGCACCT AAAACAT5TTT
251 CTATGGGAGC CAAGCCTACT GGATCCGCTG CTGCAA5ACTA TACTACTGCC
301 GTAGATGAGC CTAACCCGGC CTACAATAAG CATTTACACG ATGCAGAGTG
351 GTTCACTAAT GCAGGCTTCA TTGCCT5TAAA CATTTGGGAT CGCTTTGATG
401 TTTTCTGTAC TTTAGGAGCT TCTAAT5GGT ACATTAGAGG AAAC5TCTACA
451 GCGTTCAATC TCGTTGGT5T ATT5CGGAGT AAAGGTACTA CTGTAAATGC
501 AAATGAACTA CCAAACGTTT CTTTAAGTAA CGGAGT5GTT GAACTTTACA
551 CAGACACCTC TTTCTCT5TGG AGCGTAGGCG CTCGTGGAGC CTTATGGGAA
601 TGCGGT5GTG CAACTTTGGG AGCTGAATTC CAATATGCAC AGTCCA5AACC
651 TAAAGTTGAA GAACTTAATG TGATCTGTAA CGTATCGCAA TTCTCTGTAA
701 ACAAA5CCCA GGGCTAT5AA GGC5T5GCTT TCC5CT5TGCC AACAGACGCT
751 GGCGTAGCAA CAGCTACTGG AACAAAGTCT GCGACCATCA ATTATCATGA
801 ATGGCAAGTA GGAGCCTCTC TATCTTACAG ACTAA5ACTCT TTAGTGCCAT
851 ACATTGGAGT ACAATGGTCT CGAGCA5ACTT TTGATGCTGA TAACATCCGC
901 ATTGCTCAGC CAAA5ACTACC TACAGCTGTT TTA5ACTTAA CTGCATGGAA
951 C5CTTCTTTA CTAGGAAATG CCACAGCATT GTCTACTACT GATT5GTTCT
1001 CAGACTTCAT GCAAAT5GTT TCCTGTCAGA TCAACA5AGTT TAAATCTAGA
1051 AAAGCTTGTG GAGTTACTGT AGGAGCTACT TTAGTTGATG CTGATAAATG
1101 GTCACTTACT GCAGAAGCTC GTTTAAT5TAA CGAGAGAGCT GCTCACGTAT
1151 CTGGTCAGTT CAGATTCTAA
```

The PSORT algorithm predicts an outer membrane location (0.707).

The protein was expressed in *E.coli* and purified as a GST-fusion (Figure 30A) and as a his-tag product. The recombinant GST-fusion protein was used to immunise mice, whose sera were used in a Western blot (Figure 30B) and for FACS analysis (Figure 30C).

45 The cp6998 protein was also identified in the 2D-PAGE experiment (Cpn0695) and showed good cross-reactivity with human sera, including sera from patients with pneumonitis.

These experiments show that cp6998 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

## Example 31

The following *C.pneumoniae* protein (PID 4377102) was expressed <SEQ ID 61; cp7102>:

```

1  MKHTFTKRVL FFFFLVIP LLLNLMVVG FFSFAAKANL VQVLHTRATN
51  LSIEFEKKLT IHKLFLDRLA NTLALKSYAS PSAEPYAQAY NEMMALSNTD
101 FSLCLIDPFD GSVRTKNPGD PFIRYLKQHP EMKKKLSAAV GKAFLLTIPG
151 KPLLHYLILV EDVASWDSTT TSGLLVSFYP MSFLQKDLFQ SLHITKGNIC
201 LVNKYGEVLF CAQDSESSFV FSLDLPLNPQ FQARSPSAIE IEKASGILGG
251 ENLITVSINK KRYLGLVLNK IPIQGTYTLS LVPVSDLIQS ALKVPLNICF
301 FYVLAFLLMW WIFSKINTKL NKPLQELTFC MEAAWRGNHN VRFEPQPYGY
10  351 EFNELGNIFN CTLLLLLSNI EKADIDVHSG EKLQKELGIL SSLQSALLSP
401 DFPTFPKVT F SSQHLRRRQL SGHFNQWTVQ DGGDTLLGII GLAGDIGLPS
451 YLYALSARSL FLAYASSDVS LQKISKDTAD SFSKTTEGNE AVVAMTFIKY
501 VEKDRSLELL SLSEGAPTMF LQRGESFVRL PLETHQALQP GDRLICLTGG
551 EDILKYPSQL PIEELKDP L NPLNTENLID SLTMMLNNET EHSADGTLTI
15  601 LSFS*

```

A predicted signal peptide is highlighted.

The cp7102 nucleotide sequence <SEQ ID 62> is:

```

1  ATGAAACATA CCTTTACCAA GCGTGTCTA TTTTTTTCT TTTTAGTGAT
51  TCCCATTTCC CTACTCCTCA ATCTTATGGT CGTAGGTTTT TTCTCATTTT
20  101 CTGCCCGCTAA AGCAAAATTA GTACAGGTCC TCCATACCCG TGCTACGAAC
151 TTAAGTATAG AATTCGAAAA AAAACTGACG ATACACAAGC TTTTCCTCGA
201 TAGACTTGCC AACACATTAG CCTTAAATC CTATGCATCT CCTTCTGCAG
251 AGCCCTATGC ACAGGCATAC AATGAGATGA TGGCACTCTC CAATACAGAC
301 TTTTCCTTAT GCCTTATAGA TCCCTTTGAT GGATCTGTAA GGACGAAAAA
25  351 TCCTGGAGAC CCTTTCATTC GCTATCTAAA ACAGCATCCT GAAATGAAGA
401 AAAAGCTATC CGCAGCTGTA GGGAAAGCCT TTTTATTGAC CATTCAGGT
451 AAACCACTTT TACATATCTT TATTCTAGTT GAAGATGTCG CATCTTGGGA
501 TTCTACAACG ACTTCAGGAC TGCTTGTAAG TTTCTATCCC ATGCTCTTTT
551 TACAGAAAGA TTTATTCCAA TCCTTACACA TCACCAAGG AAATATCTGC
30  601 CTTGTAAATA AGTATGGCGA GGTCTCTTTC TGTGCTCAGG ACAGTGAATC
651 TTCTTTTGTA TTTTCTCTAG ATCTCCCTAA TTTACCGCAA TTCCAAGCAA
701 GAAGCCCTC TGCCATAGAA ATTGAGAAAG CTTCTGGAAT TCTTGGTGGG
751 GAGAACTTAA TCACAGTGAG TATCAACAAG AAACGCTACC TAGGATTGGT
801 ACTGAATAAA ATTCTATACC AAGGGACCTA CACTCTATCT TTAGTTCAG
35  851 TTTCTGATCT CATCCAATCC GCCTTGAAAG TTCCTCTCAA TATTGTGTTT
901 TTCTATGTAC TTGCTTTCCT CCTCATGTGG TGGATTTTCT CTAAGATCAA
951 CACCAAACTT AACAAGCCTC TTCAAGAACT GACCTTCTGT ATGGAAGCTG
1001 CCTGGCGAGG AAACATAAC GTGAGGTTTG AACCCAGCC TTACGGTTAT
1051 GAATTCAATG AACTAGGAAA TATTTTCAAT TGCACCTCC TACTCTTATT
40  1101 GAATTCATT GAGAAAGCAG ATATCGATTA CCATTAGGC GAAAAATTAC
1151 AAAAAGAATT AGGGATTTTA TCTTCACTAC AAAGTGCCTT ACTAAGTCCG
1201 GATTTCCCTA CGTTCCCTAA AGTTACCTTT AGTTCCCAAC ATCTCCGAG
1251 AAGGCAACTT TCCGGTCATT TTAATGGTTG GACAGTTCAA GATGGTGGCG
1301 ATACCTTTT AGGGATCATA GGGCTCGCTG GCGATATTGG TCTTCTTCC
45  1351 TATCTCTATG CTTTATCCGC ACGGAGTCTT TTTCTTGCC TATGCTTCTC
1401 GGACGTTTCG TTACAAAAAA TCAGCAAGGA TACTGCCGAC AGCTTCTCAA
1451 AAACAACAGA AGGCAATGAG GCTGTAGTTG CTATGACTTT CATTAATAT
1501 GTAGAAAAAG ATCGATCTCT AGAGCTCCTC TCGTTAAGCG AGGGAGCTCC
1551 TACCATGTTT CTACAACGAG GAGAATCTTT CGTACGCTC CCCTTAGAGA
50  1601 CTCACCAAGC TCTACAGCCT GGAGATCGGT TGATCTGCCT CACTGGAGGA
1651 GAAGACATCC TCAAGTACTT TTCTCAGCTT CCTATTGAAG AGCTCTTAAA
1701 AGATCCTTTA AACCTCTAA ATACAGAGAA TCTTATTGAT TCTCTAACCA
1751 TGATGTAAA CAACGAAACC GAACATTCTG CAGATGGAAC TCTGACCATC
1801 CTTTCATTTT CATAA

```

55 The PSORT algorithm predicts an inner membrane location (0.338).

The protein was expressed in *E.coli* and purified as a his-tag product and as a GST-fusion product. The purified GST-fusion product is shown in Figure 31A. The recombinant GST-fusion protein was used to immunise mice, whose sera were used in a Western blot and for FACS analysis (Figure 31B).

These experiments show that cp7102 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 32

The following *C.pneumoniae* protein (PID 4377106) was expressed <SEQ ID 63; cp7106>:

```

5      1  MKDLGTLGGT  SSTAKTVSPD  GKVIIMGRSQI  ADGSWHAFMC  HTDFSSNNVL
      51  FDLNNTYKTL  RENGRLNSI  FNLQNMMLQR  ASDHEFTFEG  RSNIALGAGL
     101  YVNALQNLPS  NLAAQYFGIA  YKIRPKYRLG  VFLDHNFFSS  VPNNFNVSHN
     151  RLWMGAFIGW  QSDALGSSV  KVSFGYKGKQ  ATITREQLN  TEAGSGESHF
     201  EGVAAQIEGR  YGKSLGSHV  VQPFLLGLQFV  HITRKEYTEN  AVQFPVHYDP
    10  251  IDYSTGVVYL  GIGSHIALVD  SLHVGTMRGM  EQNFAAHTDR  FSGSIASIGN
     301  FVFEKLDVTH  TRAFEAEMRV  YELPYLQSLN  LILRVNQPL  QGVMGFSSDL
     351  RYALGF*

```

The cp7106 nucleotide sequence <SEQ ID 64> is:

```

15      1  ATGAAAGATT  TGGGGACTCT  TGGGGGTACC  TCTTCTACAG  CAAAAACAGT
      51  GTCCCCAGAT  GGTAAAGTGA  TCATGGGTAG  ATCACAAATT  GCTGATGGCA
     101  GTTGGCAGCG  ATTTATGTGT  CATACGGATT  TCTCCTCTAA  TAATGTACTC
     151  TTTGATCTCG  ATAATACGTA  TAAACTCTA  AGAGAAAATG  GCCGTCAGCT
     201  AAATTCCTAC  TTCAACCTAC  AAAATATGAT  GTTACAGAGA  GCCTCAGATC
     251  ATGAGTTCAC  AGAGTTTGG  AGGAGTAACA  TCGCTCTTGG  TGCCGGGCTT
    20  301  TATGTGAATG  CCTTGCAGAA  TCTCCCTAGC  AATTTAGCAG  CACAATATTT
     351  TGGAATCGCA  TACAAAATAC  GTCCCTAAATA  TCGTTTGGGG  GTGTTTGTGG
     401  ACCATAATTT  CAGCTCCAC  GTTCCCTAATA  ATTTTAACGT  AAGCCACAAT
     451  AGACTCTGGA  TGGGAGCCTT  TATTGGATGG  CAGGATTCTG  ATGCTCTAGG
     501  ATCTAGTGTC  AAGGTGCTT  TCGGATATGG  AAAACAAAAA  GCCACGATTA
    25  551  CAAGAGAGCA  ATTAGAGAAT  ACAGAAGCCG  GGAGTGGGGA  GAGCCATTTT
     601  GAAGGGGTCG  CTGCTCAGAT  AGAAGGGCGG  TATGGTAAGA  GCCTCGGAGG
     651  ACATGTACAG  GTCCAGCCTT  TCCTAGGACT  GCAGTTTGTC  CACATTACAA
     701  GGAAAGAATA  TACCGAAAAT  GCAGTGCAAT  TTCTGTGACA  CTATGATCCT
     751  ATAGACTATT  CTACAGGTGT  AGTGTATTTA  GGAATGGAT  CTCATATTGC
    30  801  ACTTGTAGAT  TCTTTACATG  TAGGCACACG  CATGGGAATG  GAGCAAACT
     851  TTGCAGCCCA  TACGGACAGG  TTCTCAGGAT  CTATAGCGTC  TATTGGAAAC
     901  TTTGTGTTTG  AAAAGCTTGA  TGTGACTCAC  ACAAGGGCAT  TTGCGGAAAT
     951  GCGTGTCAAC  TATGAGCTTC  CCTATCTACA  GTCTCTGAAT  CTTATTCTAC
    1001  GAGTTAATCA  ACAGCCTCTA  CAAGGGGTTA  TGGGATTTTC  CAGTGATCTT
    35  1051  AGGTATGCCT  TAGGATTCTA  A

```

The PSORT algorithm predicts a cytoplasmic location (0.224).

The protein was expressed in *E.coli* and purified as a his-tag product and as a GST-fusion product. The purified GST-fusion product is shown in Figure 32A. The recombinant GST-fusion protein was used to immunise mice, whose sera were used in a Western blot (Figure 32B) and for FACS analysis (Figure 32C).

This protein also showed very good cross-reactivity with human sera, including sera from patients with pneumonitis.

These experiments show that cp7106 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### 45 Example 33

The following *C.pneumoniae* protein (PID 4377228) was expressed <SEQ ID 65; cp7228>:

```

      1  MTAVLILTSF  PSEBSARSLA  RHLITERLAS  CVHVFPKGTS  TYLWEGKLCB
     51  SEEHHIQIKS  IDIRFSEICL  AIQEFSGYEV  PEVLLFPIEN  GDPRYLNWLT
    101  ILSYPEKPPL  SD*

```



The cp7228 nucleotide sequence <SEQ ID 66> is:

```

      1  ATGACTGCTG  TTCTTATTCT  TACATCTTTC  CCTTCGGAGG  AAAGTGCTCG
    51  CTCCTTAGCT  AGACATCTGA  TTACAGAGCG  TCTTGCTTCC  TGTGTGCATG
   101  TATTCCCTAA  AGGCACATCG  ACATATCTAT  GGAAGGCAA  GCTATGTGAG
   151  TCTGAAGAAC  ATCATATACA  AATCAAATCG  ATAGACATAC  GCTTCTCGGA
   201  AATTGTCTTT  GCTATTCAGG  AGTTCTCTGG  CTATGAGGTT  CCTGAAGTCT
   251  TACTATTTCC  TATTGAAAAT  GGGGATCCGA  GGTACTTGAA  TTGGTTAACG
   301  ATTCCTCAGT  ATCCAGAGAA  GCCTCCGCTT  TCAGATTAG

```

The PSORT algorithm predicts an inner membrane location (0.040).

- 10 The protein was expressed in *E.coli* and purified as a his-tag product and as a GST-fusion product, as shown in Figure 33A (his-tag = left-hand arrow, GST = right-hand arrow). The proteins were used to immunise mice, whose sera were used in a Western blot (Figure 33B) and FACS analysis.

These experiments show that cp7228 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### 15 Example 34

The following *C.pneumoniae* protein (PID 4377170) was expressed <SEQ ID 67; cp7170>:

```

      1  MNSKMLKHLR  LATLSFSMFF  GIVSSPAVYA  LGAGNPAAPV  LPGVNPEQTG
    51  WCAPQLCNSY  DLFAALAGSL  KFGFYGDYVF  SESAHITNVP  VITSVTTSQT
   101  GTTPITITST  KNVDFDLNNS  SISSSCVFAT  IALQETSPAA  IPLLDIAFTA
   201  RVGGLKQYYR  LPLNAYRDFE  SNPLNAESEV  TDGLIEVQSD  YGIVWGLSLQ
   251  KVLWKDGVSF  VGVSAHYRHG  SSPINYLIIVY  NKANPEIYFD  ATDGNLSYKE
   301  WSASIGISTY  LNDYVLPYAS  VSIGNPSRKA  PSDSFTLELE  QFTNPFKFKIR
      301  KITNFDVRNF  CFGTTCCISN  NFYYSVEGRW  GYQRAINITS  GLQF*

```

A predicted signal peptide is highlighted.

- 25 The cp7170 nucleotide sequence <SEQ ID 68> is:

```

      1  ATGAATAGCA  AGATGCTAAA  ACATTACGTT  TTAGCAACCC  TTTCCTTCTC
    51  TATGTTCTTC  GGGATTGTAT  CTTCTCCCGC  AGTATATGCC  CTAGGGGCTG
   101  GAAACCCCTG  AGCTCCAGTA  CTCCCAGGTG  TGAATCCTGA  GCAAACGGGA
   151  TGGTGTGCCT  TCCAACCTTG  TAATAGTTAC  GATCTTTTTC  CTGCTCTTGC
   201  AGGAAGCCTC  AAATTGGGT  TCTATGGAGA  TTATGCTCTC  TCAGAAAGTG
   251  CCCATATTAC  CAATGTCCCT  GTCATTACCT  CCGTTACGAC  TTCAGGCACA
   301  GGAACAACGC  CAACCATTAC  CTCTACAAC  AAAAACGTA  ACTTTGATCT
   351  TAACAACAGC  TCCATCAGCT  CGAGCTGTGT  TTTTGCAACC  ATAGCTCTAC
   401  AGGAACATC  CCCAGCTGCC  ATTCCCTTT  TAGATATAGC  CTTCACTGCA
   451  CGTGTCCGAG  GACTTAAGCA  GTACTACCGC  CTCCCTCTCA  ATGCTTACAG
   501  AGACTTCACT  TCAAATCCTT  TAAATGCAGA  ATCTGAAGTT  ACAGATGGTC
   551  TCATTGAAGT  CCAGTCAGAC  TATGGAATTG  TCTGGGGTCT  GAGTTTACAA
   601  AAAGTATTGT  GGAAGATGG  AGTGTCTTTT  GTAGGGGTGA  GCGCTGACTA
   651  CCGTCACGGT  TCCAGTCCCA  TCAACTATAT  CATCGTTTAC  AACAAGGCCA
   701  ACCCCGAGAT  CTATTTCGAT  GCTACTGATG  GAAACCTAAG  CTATAAAGAA
   751  TGGTCTGCAA  GCATCGGCAT  CTCTACGTAT  CTTAATGACT  ATGTGCTTCC
   801  CTATGCATCC  GTATCTATAG  GAAATACTTC  AAGAAAAGCT  CCTTCTGATA
   851  GCTTTCACAG  ACTCGAAAAG  CAATTACGA  ATTTTAAAT  TAAAATTCGT
   901  AAAATCACAA  ACTTCGACAG  AGTAACTTC  TGCTTCGGAA  CTACCTGCTG
   951  CATCTCAAAT  AACTTCTACT  ATAGTGTAGA  AGGCCGTTGG  GGATATCAGC
  1001  GTGCTATCAA  CATTACGTCA  GGTCTGCAGT  TTTAG

```

The PSORT algorithm predicts a bacterial outer membrane location (0.936).

The protein was expressed in *E.coli* and purified as a his-tag product and as a GST-fusion product.

The purified GST-fusion product is shown in Figure 34A. The GST-fusion protein was used to

- 50 immunise mice, whose sera were used in a Western blot (34B) and for FACS analysis (34C).

The cp7170 protein was also identified in the 2D-PAGE experiment (Cpn0854).

These experiments show that cp7170 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 35

5 The following *C.pneumoniae* protein (PID 4377072) was expressed <SEQ ID 69; cp7072>:

```

1  MDIKKLFCLF LCSSLIAMSP IYKKTGDYK LTLTGINIID RNGLSETICS
51  KEKLLKKYTKV DFLAPQPYQK VMRMKKNKRG DNVSCLTAYH TNGQIKQYLE
101 CLNNRAYGRY REWHVNGNIK IQAEVIGGIA DLHPSAESGW LFDQTTTFAYN
151 DEGILEAAIV YEKGLLEGSS VYYHTNGNIW KECFYHKGVP QGKFLTYTSS
10  201 GKLLKEQNYQ QGKRHGLSIR YSEDSEEDVL AWEYHHEGRL LKAELYDPQT
251 HEIYATHEG NGIQAIYGKY AVIETRAFYR GEPYGVTRF DNSGTQIVQT
301 YNLLQGAHKG BEFFFPYPETG KPKLLLNWHE GILNGIVKTW YPGGTLESCK
351 ELVNNKKSGL LTIYYPEGQI MATEEYDNDL LIKGEYFRPG DRHPYSKIDR
401 GCGTAVFFSS AGTITTKIPY QDGKPLLN*

```

15 A predicted signal peptide is highlighted.

The cp7072 nucleotide sequence <SEQ ID 70> is:

```

1  ATGGATATAA AAAAAGCTCTT TTGCTTATTT CTATGTTCTT CTCTAATTGC
51  CATGAGTCCC ATTTATGGGA AAACAGGTGA CTATGAGAAA CTCACCCCTA
101  CAGGGATCAA TATCATTTGAT AGAAACGGCC TGTGAGAAAC TATTTGCTCT
20  151 AAAGAGAAGC TAAAGAAATA CACCAAGGTA GACTTTCTTG CCCCCAGCC
201  CTATCAAAAG GTCATGAGGA TGTATAAAAA CAAACGCGGA GATAACGTTT
251  CTTGTTTAAC AGCCTATCAC ACTAACGGGC AAATTAAGCA GTACCTGGAG
301  TGTCTCAATA ATCGTGCTTA TGAAGATAT CGTGAATGGC ACGTCAACGG
351  GAATATCAAA ATCCAAGCTG AGGTTATCGG AGGTATGCG GATCTTCATC
25  401 CCTCAGCAGA GTCTGGCTGG CTATTTGATC AAACACATT TGCCTATAAT
451  GATGAAGGTA TCTTAGAAGC CGCTATCGTC TATGAAAAAG GGCTGCTCGA
501  AGGATCTTCG GTGTATTACC ATACTAATGG GAATATTGG AAAGAGTGTC
551  CCTATCATAA GGGAGTTCTT CAAGGTAAAT TCCTGACATA CACATCTTCG
601  GGGAAACTGC TCAAAGAACA GAATTACCAA CAAGGCAAAA GACACGGTCT
30  651 TTCGATTTCG TACAGCGAAG ATTCCGAAGA AGATGTTTAA GCCTGGGAAG
701  AATATCATGA GGGACGACTC CTAAAAGCAG AGTACTTAGA TCCTCAAAC
751  CACGAAATCT ATGCGACTAT ACACGAAGGG AACGGCATTC AAGCAATCTA
801  CCGCAAGTAT GCCGTTATAG AAACAGGGC ATTTTACCGA GGGGAACCTT
851  ATGGAAGAGT TACCAGATTG GACAACTCCG GAACACAGAT TGTCCAAACG
35  901 TATAACCTTT TGCAAGGCGC GAAGCACGGA GAAGAATTTT TCTTTTATCC
951  TGAGACAGGG AAACCCAAAG TGCTTCTTAA TTGGCATGAA GGAATTTTAA
1001 ATGGGATAGT AAAAAGTTGG TATCCCGGAG GAACCTTAGA AAGTTGTAAA
1051 GAACTCGTAA ATAACAAAAA ATCCGGGTGA CTGACCATTT ACTACCCTGA
1101 AGGACAGATC ATGGCGACCG AAGAGTATGA TAATGATCTT CTAATTAAG
40  1151 GAGAGTACTT CCGCCCTGGA GACCGTCATC CTTACTCTAA AATAGATCGT
1201 GGTTCGTTGA CTGCAGTATT TTTCTCGTCG GCGGGAACTA TTACTAAAAA
1251 AATCCCTTAT CAGGACGGCA AACCTTTGCT CAACCTAG

```

The PSORT algorithm predicts a periplasmic location (0.688).

45 The protein was expressed in *E.coli* and purified as a his-tag product (Figure 35A) and as a GST-fusion product (Figure 35B). The recombinant his-tag protein was used to immunise mice, whose sera were used in a Western blot (Figure 35C) and for FACS analysis.

These experiments show that cp7072 is a useful immunogen. These properties are not evident from the sequence alone.

### Example 36

50 The following *C.pneumoniae* protein (PID 4376879) was expressed <SEQ ID 71; cp6879>:

```

1  MATPAQKSPT FQDPSFVREL GSNHPVFSPL TLEERGEMAI ARVQQCGWNH
51  TIVKVSILIL ALLTILGGGL LVGLLPVAVPM FIGTGLIALG AVIFALALIL
101 CLYDSQGLPE ELPPVPEPQQ IQIEDLRNET REVLEGTLLLE VLLKDRDAKD
151 PAVPQVVVDC EKRLGMLDRK LRREEEILYR STAHLKDEER YEFLLELLEM
201 RSLVADRLEF NRRSYERFVQ GIMTVRSEEG EKEISRLQDL ISLQQQTVQD
251 LRSRIDDEQK RCWTALQRIN QSQKDIQRAH DREASQRACE GTEMDCAERQ
301 QLEKDLRRQL KSMQEWIEMR GTIHQQEKAW RKQNAKLERL QEDLRLTGIA
351 FDEQSLFYRE YKEKYSQKL DMQKILQEVN AEKSEKACLE SLVHDYQKQL
401 EQKDANLKA AAVWEEELGK QQQEDYEQTO EIRRLSTFIL EYQDSLREAE
451 KVEKDFQELQ QRYSLRQEEK QVKEKILEES MNHFADLFK AQKENMAYKK
501 KLADLEGAAA PTEIGEDDDW VLTDSASLSQ KKIRELVEEN QELLKALAFK
551 SNELTQLVAD AVEAEKEISK LREHIKEQKB GLRALDKMHA QAIKDCAAQ
601 RKCCDLESLL SPVREDAGMR FELEVELQRL QEENALRAE VERLEQEQFO
651 G*

```

The cp6879 nucleotide sequence <SEQ ID 72> is:

```

1  ATGGCAACAC CCGCTCAAAA ATCCCTACA TTTCAAGATC CTAGTTTGT
51  AAGAGAGCTA GGCAGTAACC ACCCTGTCTT TTCCCGCTA ACGCTTGAGG
101 AAAGAGGGGA GATGGCAATA GCTCGAGTCC AGCAGTGTGG ATGGAATCAT
151 ACAATTGTTA AGGTAAGTCT TATTATCTT GCTCTCTTA CTATTTTAGG
201 GGGAGGATTA CTCGTAGGAT TGCTGCCAGC AGTTCCTATG TTTATTGGAA
251 CAGGTCGTAT TGCTTTGGGA GCCGTATAT TTGCTTTGGC TTTGATTTTA
301 TGTCTTTATG ATTCTCAGGG CCTTCCTGAG GAACCTCCCT CCGTTCCTGA
351 ACCACAACA ATTACAGATT AAGATTTAAG AAACGAGACC AGAGAAGTTC
401 TTGAAGGGAC TCTTTTAGAG GTTCTCTTAA AGGATAGAGA CGCTAAGGAC
451 CCTGCGGTGC CCCAGGTGGT TGTAAGTGT GAAAAGCGTC TTGGAATGTT
501 GGATCGTAAG CTGCGACGTG AAGAGGAGAT TCTGTATCGC TCGACGGCCC
551 ATCTTAAAGA CGAGGAAAGG TATGAGTCT TGCTGGAGCT CTTGGAATG
601 CGTAGTCTGG TTGCCGATCG GCTAGAATT AACCCTAGAA GTTATGAGCG
651 ATTTGTTCAA GGAATTATGA CAGTTAGATC AGAGGAGGGG GAAAAAGAGA
701 TTTCTCGTCT ACAAGATCTA ATCAGTTTGC AGCAGCAGAC GGTGCAAGAT
751 TTAAGGAGTC GGATCGATGA CGAGCAGAAG AGATGCTGGA CGGCTTTACA
801 ACGTATTAAC CAATCTCAGA AGGATATACA ACGGGCTCAT GATCGCGAGG
851 CTTCCGAGCG TGCCGTGTGAG GGCACAGAGA TGGATTGTGC AGAACGCCAG
901 CAACTGGAGA AGGATTTAAG GAGACAGCTG AAATCTATGC AGGAGTGGAT
951 TGAGATGAGG GGCACAATCC ATCAACAAGA GAAGGCTTGG CGTAAGCAGA
1001 ATGCCAAATT AGAAAGATTA CAAGAGGATC TGAGACTTAC TGGGATGTCT
1051 TTTGACGAAC AATCTCTGTT CTATCGCGAA TATAAAGAGA AATATCTGAG
1101 TCAGAACTA GATATGCAA AGATTTTACA GGAAGTCAAC GCAGAGAAAA
1151 GTGAGAAGGC TTGCTTAGAG AGTCTGGTCC ATGACTATGA GAAGCAGCTC
1201 GAACAAAAAG ATGCTAATCT GAAGAAAGCA GCAGCTGTTT GGAAGAAGA
1251 ATTAGGGAAG CAGCAACAGG AAGACTACGA ACAAACCCAA GAAATTAGAC
1301 GTCTGAGTAC ATTCAATCTT GAGTACCAGG ACAGTCTGCG TGAGGCAGAA
1351 AAAGTTGAGA AAGATTTCCT AGAGCTACAA CAAAGGTATA GCCGTCTTCA
1401 AGAGGAGAAA CAGGTAAGG AAAAAATCTT AGAAGAAAGT ATGAATCATT
1451 TTGCCGATCT CTTTGAGAAG GCTCAAAAGG AAAACATGGC CTACAAGAAG
1501 AAGTTAGCGG ATTTAGAGGG TGCCGCTGCT CCTACTGAGA TCGGTGAGGA
1551 CGATGACTGG GACTCACAG ATTCTGCTTC TCTCAGCCAG AAGAAGATCC
1601 GCGAACTCGT GGAAGAGAAT CAAGAATCC TGAAAGCACT TGCATTFAAA
1651 TCTAACGAAT TGACTCAACT GGTGCCGAT GCTGTAGAAG CTGAAAAAGA
1701 AATCAGCAAG CTTTCGGAAC ACATAGAAGA GCAGAAAGAA GGATTACGAG
1751 CTCTTGATAA GATGCATGCA CAAGCGATCA AAGATTGCGA AGCTGCTCAG
1801 AGAAAATGCT GTGACCTTGA GAGCCTTCTC TCTCCTGTTT GAGAAGATGC
1851 TGGAATGAGA TTTGAGCTAG AGGTCGAGCT TCAAAGATTG CAAGAAGAAA
1901 ATGCACAGCT TAGAGCGGAG GTTGAAAGAC TAGAGCAAGA GCAATTTCAA
1951 GGATAA

```

The PSORT algorithm predicts an inner membrane location (0.646).

The protein was expressed in *E. coli* and purified as a his-tag product and as a GST-fusion product. The purified GST-fusion product is shown in Figure 36A. The recombinant GST-fusion protein was used to immunise mice, whose sera were used in a Western blot (Figure 36B) and for FACS analysis.

These experiments show that cp6879 is useful immunogen. These properties are not evident from the sequence alone.

**Example 37**

The following *C.pneumoniae* protein (PID 4376767) was expressed <SEQ ID 73; cp6767>:

```

1  MIKQIGRFFR AFIFIMPLSL TSCSKIDRN RIWIVGTNAT YPPFEYVDAQ
5  51  GEVVGFDIDL AKAISEKLGK QLEVREPAFD ALILNLKKHR IDAILAGMSI
101 TPSRQKEIAL LPYYGDEVQE LMVVSQRSLE TPVLPLTQYS SVAVQTGTFQ
151 EHYLLSQPGI CVRSFDSTLE VIMEVRYGKS PVAVLEPSVG RVVLKDFPNL
201 VATRLELPPE CWVLGCGLGV AKDRPEEIQT IQQAITDLKS EGVISLTKK
251 WQLSEVAYE*

```

The cp6767 nucleotide sequence <SEQ ID 74> is:

```

10 1  ATGATAAAAC AAATAGGCCG TTTTTTTAGA GCATTTATTT TTATAATGCC
51  TTTATCTTTA ACAAGTTGTG AGTCTAAAAT CGATCGAAAT CGCATCTGGA
101 TTGTAGGTAC GAATGCTACA TATCCTCCTT TTGAGTATGT GGATGCTCAG
151 GGGGAAGTTG TAGGTTTCGA TATAGATTTC GCAAAGGCAA TTAGTGAAAA
201 ACTTGGCAAG CAATTGGAAG TTAGAGAATT CGCTTTCGAT GCTTTAATTT
15 251 TAAATTTAAA AAAACATCGT ATCGATGCAA TTTTAGCAGG AATGTCCATT
301 ACTCCTTCGC GTCAGAAGGA AATCGCCTG CTTCCTTATT ATGGCGATGA
351 GGTTCAGAG CTGATGGTGG TTCTAAGCG GTCTTTAGAG ACCCCTGTGC
401 TCCCCTAAC ACAGTATTCT TCTGTTGCTG TTCAGACAGG AACGTTTCAG
451 GAGCATTTATC TTTTATCTCA GCCCGGAATT TGTGTCCGTT CTTTGATAG
20 501 CACCTTGGAG GTGATTATGG AAGTTCGTTA TGGGAAATCT CCGGTTGCCG
551 TTCTAGAACC CTCGGTAGGA CGTGTCTGTC TTAAAGACTT CCCTAATCTT
601 GTTGCAACAA GATTAGAGCT CCCTCCTGAA TGTGGGTGT TGGGCTGTGG
651 TCTCGCGGTA GCTAAAGATC GTCCTGAAGA AATACAAACG ATTCAACAAG
701 CGATTACAGA TTAAAGAGC GAAGGGGTGA TTCAATCTTT AACCAAGAAA
25 751 TGGCAACTTT CTGAAGTTGC TTACGAATAG

```

The PSORT algorithm predicts an inner membrane location (0.083).

The protein was expressed in *E.coli* and purified as a his-tag product and as a GST-fusion product. The purified his-tag product is shown in Figure 37A. The recombinant his-tag protein was used to immunise mice, whose sera were used in a Western blot (Figure 37B) and for FACS analysis (Figure 37C). The GST-fusion was also used in a Western blot (Figure 37D).

The cp6767 protein was also identified in the 2D-PAGE experiment and showed good cross-reactivity with human sera, including sera from patients with pneumonitis.

These experiments show that cp6767 is a useful immunogen. These properties are not evident from the sequence alone.

**Example 38**

The following *C.pneumoniae* protein (PID 4376717) was expressed <SEQ ID 75; cp6717>:

```

1  MMSRLRFRLA ALGIFPILLV PNSVSAKTIV ASDKREKVGVL VYDNSVEAFQ
51  QILDICIDHAN FYVELCPCMT GGRTLKEMVD HLEARMDLVP ELCSYIIIQP
101 TPTDAEDQKL LKALKERHPN RFFYVFTGCP PSTSILAPNV IEMHIKLSII
40 151 DGKYCILGGT NFEFMCTPG DEVPEKVDNP RLFVSGVRRP LAFRDQDIDL
201 RSTAFGLQLR EBYHKQFAMW DYYAHMWFI DNPEQFAGAC PPLTLEQAE
251 TVFPGFDKHE DLVLVDSSKI RIVLGGPHDK QPNPVTQEYL KLIQGARSSV
301 KLAHMYFIPK DELLNALVDV SHNHGVHLSL ITNGCHELSP AITGPHYAWN
351 RINYFALLYG KRYPLWKKWF CEKLPYERV SIYEFAIWET QLHKCMIID
45 401 DEIFVIGSYN FGKKSDAFDY ESIVVIESPE VAAKANKVFN KDIGLSIPVS
451 HGDIFSWYFH SVHHTLGHLLQ LTYMPA*

```

A predicted signal peptide is highlighted.

The cp6717 nucleotide sequence <SEQ ID 76> is:

-78-

```

1  ATGATGAGTC GGTTCGCTTT TCGCTTGGCA GCTCTTGGAA TATTTTITAT
51  TTTGCTGGTT CCTAATTCCTG TTTCAGCAAA GACAATCGTA GCTTCAGACA
101 AGGAGAAGGT TGGAGTTCCTT GTTTATGACA ATAGTGTAGA GGCCTTTCAA
151 CAGATATTGG ATTGCATAGA TCATGCAAAT TTTTATGTAG AACTGTGTCC
5  201 CTGCATGACA GGAGGCCGAA CGCTTAAAGA GATGGTAGAT CACCTCGAGG
251 CTCGTATGGA TCTGGTTCCA GAGCTCTGTA GCTATATCAT TATCCAACCC
301 ACGTTTACCG ATGCTGAAGA CAAAAATTA CTCAAAGCTC TCAAAGAACG
351 TCATCCCAAC CGGTTTTCTT ACGTTTCTTAC AGGGTGCCCA CCCTCAACAA
401 GCATCCTCGC TCCTAATGTC ATTGAAATGC ATATCAAAC TTTCTATCAT
10 451 GATGGGAAAT ATTGTATTTT AGGTGGTACC AATTTTGAAG AGTTTATGTG
501 CACTCCAGGG GATGAGGTTT CTGAGAAAGT GGATAACCCA CGTTTATTTG
551 TCAGTGGAGT GCGTCGGCCC CTAGCATTTT GTGATCAGGA TATCATGTTG
601 CGTTCTACAG CATTCGGTTT GCAGCTCAGA GAAGAATATC ATAAGCAATT
15 651 TGCTATGTGG GACTACTATG CACATCATAT GTGGTTCATT GATAATCCTG
701 AACAGTTTGC AGGCGCCTGT CCTCCACTGA CTTTAGAACA AGCCGAGGAG
751 ACAGTATTTT CTGGATTGTA CAAACATGAA GATCTTGTTT TTGTCGACTC
801 TTCCAAGATC AGGATAGTTT TAGGTGGTCC CCACGATAAG CAACCCAATC
851 CTGTGACTCA AGAATATTTG AAACCTATCC AGGGAGCTAG ATCTTCTGTG
901 AAGCTTGCTC ACATGTATTT CATCCCTAAG GACGAGCTTT TAAATGCTCT
20 951 TGTCGACGTT TCTCATAATC ACGGTGTTCA TCTGAGTTTA ATTACGAACG
1001 GCTGTCATGA ATTAAGTCCT GCAATTACAG GACCCATATG TTGGGGAAAC
1051 CGTATTAACT ATTTTCGCCTT GCTCTATGGG AAACGGTATC CTCTTTGGAA
1101 AAAATGGTTT TGCGAAAAGC TAAAACCTTA TGAGCGGGTT TCTATTATG
1151 AGTTTGCTAT TTGGGAAACG CAGTTGCACA AGAAGTGAT GATTATCGAT
25 1201 GATGAAATTT TTGTGATCGG AAGTTATAAT TTTGGAAAGA AAAGTGATGC
1251 CTTTGATTAC GAAAGTATTG TAGTTATCGA ATCTCCAGAA GTCGCTGCAA
1301 AAGCTAACAA AGTCTTCAAT AAAGATATCG GATTGTGAT TCCTGTAAGT
1351 CATGGCGACA TTTTCTCTTG GTATTTCAT TCCGTACACC ACACTTGGG
1401 ACATTTGCAG CTGACCTATA TGCCAGCCTA G

```

The PSORT algorithm predicts a periplasmic location (0.939).

The protein was expressed in *E.coli* and purified as a GST-fusion (Figure 38A), as a his-tagged protein, and as a GST/his fusion product. The proteins were used to immunise mice, whose sera were used in a Western blot (Figure 38B) and for FACS analysis.

These experiments show that cp6717 is a useful immunogen. These properties are not evident from the sequence alone.

### Example 39

The following *C.pneumoniae* protein (PID 4376577) was expressed <SEQ ID 77; cp6577>:

```

1  MKRLLPSTFL LVLGSTSAAH ANLGYVNLKR CLRESDLGKK ETEELRAMKQ
51  QFVKNAEKIE BELTSYNNKL QDEDYMESLS DSASEELRKK FEDLSGEYNA
40 101 YQSQQYQSIN QSNVKRIQKL IQEVKIAAES VRSKEKLEAI LNEEAVLAIA
151 PGTDRTTEII AILNESFKKQ N*

```

A predicted signal peptide is highlighted.

The cp6577 nucleotide sequence <SEQ ID 78> is:

```

1  ATGAAAAAAT TATTATTTTC TACATTTCCTT CTTGTTTTAG GATCAACAAG
45 51  CGCAGCTCAT GCAAATTTAG GCTATGTTAA TTTAAAGCGA TGTCTTGAAG
101 AATCCGATCT AGGTAAAAAG GAAACTGAAG AATTGGAAGC TATGAAACAG
151 CAGTTTGTA AATGCTGTA GAAATAGAA GAAGAACTCA CTTCTATTTA
201 TAATAAGTTG CAAGATGAAG ATTACATGGA AAGCCTATCG GATTCTGCCT
251 CTGAAGAGTT GCGAAAGAAA TTCGAAGATC TTTGAGGAGA GTACAATGCG
50 301 TACCAGTCTC AGTACTATCA ATCTATCAAT CAAAGTAATG TAAAACGCAT
351 TCAAAAACCT ATTCAAGAAG TAAAAATAGC TGCAGAAATC GTGCGGTCCA
401 AAGAAAAACT AGAAGCTATC CTTAATGAAG AAGCTGTCTT AGCAATAGCA
451 CCTGGGACTG ATAAAACAAC CGAAATTATT GCTATTCTTA ACGAATCTTT
501 CAAAAAACAA AACTAG

```

The PSORT algorithm predicts a periplasmic space location (0.932).

The protein was expressed in *E.coli* and purified as a his-tag product (Figure 39A) and as a GST-fusion product (Figure 39B). The recombinant GST-fusion protein was used to immunise mice, whose sera were used in a Western blot (Figure 39C) and for FACS analysis.

The cp6577 protein was also identified in the 2D-PAGE experiment.

- 5 These experiments show that cp6577 is a useful immunogen. These properties are not evident from the sequence alone.

#### Example 40

The following *C.pneumoniae* protein (PID 4376446) was expressed <SEQ ID 79; cp6446>:

```

10      1  MKQPMSLIFS SVCLGLGLGS LSSCNQKPSW NYHNTSTSEE FFVHGKNSVS
      51  QLPHPYSAFR TTQIFSEEHN DPYVVAKTDE ESRKIWREIH KNLKIKGSYI
      101 PISTYGSLMH PKSAALTLKT YRPHPIWING YERSFNIDTG KYLKNGSRRR
      151 TSHDGPKNRA VLNLKSSGR RCNAIGLEMT EEDFVIARRR EGVYSLYPVE
      201 VCSYPQGNPF VIAYAWIADE SACSKEVLPV KGYYSLVWES VSSSDSLNAF
      251 GDSFAEDYLR STFLANGTSI LCVHESYKKV PPQP*
```

- 15 A predicted signal peptide is highlighted.

The cp6446 nucleotide sequence <SEQ ID 80> is:

```

20      1  ATGAAACAGC CCATGCTCTCT TATCTTTTCA AGTGTATGTT TAGGATTAGG
      51  TCTTGGATCT CTTCCTCCTCT GTAATCAAAA GCCCTCTTGG AATTATCACA
      101  ACACTTCAAC GAGCGAAGAA TTCTTTGTTC ATGGAAATAA GAGTGTTCG
      151  CAACTGCTCT ATTATCCTTC TGCATTTCTG ACGACTCAAA TCCTTTCTGA
      201  AGAGCACAAT GATCCTTATG TCGTAGCTAA GACTGATGAA GAGTCTCGTA
      251  AAATTTCGAG AGAAATCCAT AAAAATCTCA AAATCAAAGG TTCTTACATT
      301  CCCATATCGA CTTATGGAAG TCTGATGCAC CCAAATCAG CAGCTCTTAC
      351  ATTAAAAAACG TATCGTCCAC ATCCTATTTG GATAAATGGA TACGAGCGTT
      401  CTTTTAATAT AGACACAGGA AAGTACTTAA AAAACGGAAG TCGCCGTAGA
      451  ACTTCTCAGC ATGGTCCGAA AAATCGAGCT GACTGAATC TCATTAAATC
      501  TTCGGGACGA CGCTGTAATG CTATAGGCCT TGAGATGACA GAAGAAGACT
      551  TTGTAATAGC TAGAAGCCGA GAAGTGTTT ATAGCCTGTA TCCCGTTGAA
      601  GTGTGCTCGT ATCCTCAGGG GAATCCTTTT GTCATGCTT ATGCCTGGAT
      651  TGCAGATGAG AGTGTCTGCT CAAAAGAGGT CCTACCTGTA AAAGGGTACT
      701  ATTCCTTAGT CTGGGAAAGC GTTCTCTCCT CTGATCTCTT GAATGCTTTT
      751  GGAGATTCCT TTGCAGAGGA CTACCTCAGA AGCAGGTTTT TAGCAAACGG
      801  AACTTCTATA CTCGTGTGTC ATGAAAGCTA TAAGAAAGTT CCTCCTCAGC
      851  CCTAA
```

- 35 The PSORT algorithm predicts an inner membrane location (0.177).

The protein was expressed in *E.coli* and purified as a his-tag product and a GST-fusion product. The GST-fusion product is shown in Figure 40A. The recombinant his-tag protein was used to immunise mice, whose sera were used in a Western blot (Figure 40B) and for FACS analysis.

- 40 These experiments show that cp6446 is a useful immunogen. These properties are not evident from the sequence alone.

#### Example 41

The following *C.pneumoniae* protein (PID 4377108) was expressed <SEQ ID 81; cp7108>:

```

45      1  MSKKIKVLGH LTLCTLFRGV LCAAALSNIG YASTSQESPY QKSIEDWKGY
      51  TFTDLELLSK EGWSEAHAVS GNGSRIVGAS GAGQGSVTAV IWESHLIKHL
      101  GTLGGREASSA EGISKDGEVV VGWSDTREGY THAFVFDGRD MKDLGTLGAT
      151  YSVARGVSGD GSIIVGVSAT ARGEDYGWQV GVKWERGKIK QLKLLPQGLW
```

-80-

201 SEANAISEDG TVIVGRGEIS RNHIVAVKWN KNAVYSLGTL GGSVASAEAI  
 251 SANGKVIVGW STTNNGETHA FMHKDETMHD LGTLGGGFSV ATGVSADGRA  
 301 IVGFSAVKTG EIHAFYYAEG EMEDLTTLGG EEARVFDISS EGNDIIGSIK  
 351 TDAGAERAYL FHIHK\*

5 A predicted signal peptide is highlighted.

The cp7108 nucleotide sequence <SEQ ID 82> is:

1 ATGAGTAAGA AGATAAAGGT TCTAGGTCAT TTGACGCTCT GCACTCTGTT  
 51 TAGAGGAGTG CTGTGTGCAG CGGCCCTTTC CAACATAGGA TATGCGAGTA  
 101 CTTCTCAGGA ATCACCATAT CAGAAGTCTA TAGAAGACTG GAAAGGGTAT  
 151 ACCTTTACAG ATCTTGAGTT ACTGAGTAAG GAAGGGTGGT CTGAAGCTCA  
 201 TGCAGTTTCT GGAAATGGCA GTAGAATTGT AGGAGCTTCG GGAGCTGGCC  
 251 AAGGTAGTGT GACTGCTGTC ATATGGGAAA GTCACCTGAT AAAACATCTC  
 301 GGCACTTTAG GTGGCGAGGC TFCATCTGCA GAGGGAATT CAAAGGATGG  
 351 AGAGGTGGTC GTTGGGTGGT CAGATACTAG AGAGGGATAT ACTCATGCCT  
 401 TTGCTTTCGA CGGTAGAGAT ATGAAAGATC TCGGTACTCT AGGAGCTACC  
 451 TATTCTGTAG CAAGGGGTGT TTCTGGAGAT GGTAGTATCA TCGTAGGAGT  
 501 CTCTGCAACT GCTCGTGGAG AGGATTACGG ATGGCAAGTT GGTGTCAAGT  
 551 GGGAAAAGG GAAAATCAAA CAATTGAAGT TGTTCCTCA AGGTCTCTGG  
 601 TCTGAGGCGA ATGCAATCTC TGAGGATGGT ACGGTGATTG TCGGGAGAGG  
 651 TAGGAGGAGG TTTTCTGTG CCAATCACA TCGTTGCTGT AAAATGGAAT AAAAATGCTG  
 701 TGTATAGTTT GGGGACTCTC GGAGGTAGTG TCGCTTCAGC AGAGGCTATA  
 751 TCGGCAAATG GGAAAGTAAT TGTAGGATGG TCCACGACTA ATAATGGTGA  
 801 GACTCATGCC TTTATGCACA AAGATGAGAC AATGCACGAT CTCGGCACTC  
 851 TAGGAGGAGG TTTTCTGTG CCAACTGGAG TTCTGTCTGA TGGGAGAGCC  
 901 ATCGTAGGAT TTTCAGCAGT GAAGACCGGA GAAATTCATG CTTTTACTA  
 951 TGCAGAAGGA GAAATGGAGG ATTTAACAAC TTGGGAGGG GAAGAAGCTC  
 1001 GAGTGTTCGA CATATCTAGC GAAGGAAACG ATATCATTGG CTCTATAAAA  
 1051 ACTGACGCTG GAGCTGAACG CGCCTATCTG TTCCATATAC ATAAATAA

The PSORT algorithm predicts an outer membrane location (0.921).

30 The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 41A. The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 41B) and for FACS analysis (Figure 41C). A his-tagged protein was also expressed.

The cp7108 protein was also identified in the 2D-PAGE experiment.

These experiments show that cp7108 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

#### Example 42

The following *C.pneumoniae* protein (PID 4377287) was expressed <SEQ ID 83; cp7287>:

1 MVAKRTVRSY RSSFHSVTV AILSAGIAFE AHSLHSSELD LGVFNKQFRE  
 51 HSAHVEEAQT SVLKGSDFVN PSQKESEKVL YTQVPLTQGS SGESLDLADA  
 101 NFLEHFQHLF EETTTFGIDQ KLVWSDLDTR NFSQPTQEPD TSNAVSEKIS  
 151 SDTKENRKDL ETEDPSKKSG LKEVSSDLPK SPETAVAAIS EDLEISENIS  
 201 ARDPLQGLAF FYKNTSSQSI SEKDSSFQGI IFSGSGANSF LGFENLKAPK  
 251 SGAAVYSDRD IVFENLVKGL SFISCESLED GSAAGVNIVV THCGDVT LTD  
 301 CATGLDLEAL RLVKDFSRGG AVFTARNHEV QNNLAGGILS VVGKGAIVV  
 351 EKNSAEKSNF GAFACGSFVY SNNENTALWK ENQALSGGAI SSASDIDIQG  
 401 NCSAIEPSGN QSLIALGEHI GLTDFVGGGA LAAQGTTLR NNAVVCVKV  
 451 TSKTHGAIL AGTVDLNETI SEVAFKQNTA ALTGGALSAN DKVIIANNFG  
 501 EILFEQNEVR NHGGAICYGC RSNPKLEQKD SGENINIIGN SGAITFLKNK  
 551 ASVLEVMTQA EDYAGGALW GHNVLLDSNS GNIQFIGNIG GSTFWIGEVV  
 601 GGGAILSTDR VTISNNSGDV VFKGNKGQCL AQKYVAPQET APVESDASST  
 651 NKDEKSLNAC SHGDHYPPKT VEEVPPSLI EEHPVVSSTD IRGGGAILAQ  
 701 HIFITDNFTN LRFSGNLGGG EESSTVGDLA IVGGGALLST NEVNVCSNQ  
 751 VVFSNDVTSN GCDSGGAILA KKVDISANHS VEFVSNSSGK FGGAVCALNE  
 801 SVNITDNGSA VSFSKNRTRL GGAGVAAPQG SVTICGNQGN IAFKENFVFG

-81-

5 851 SENQSRGGGA IANSSSVNIQ DNAGDILFVS NSTGSYGGAI FVGSVLVASEG  
 901 SNPRTLTITG NSGDILFAKN STQTAASLSE KDSFGGGAIY TQNLKIVKNA  
 951 GNVSPFYGNRA PSGAGVQIAD GGTVCLEAFG GDILFEGNIN FDGSFNAILH  
 1001 CGNDSKIVEL SAVQDKNIIF QDAITYEENT IRGLPDKDVS PLSAPSLIFN  
 1051 SKPQDDSAQH HEGTIRFSRG VSKIPQIAAI QEGTLALSON AELWLAGLQK  
 1101 ETGSSIVLSA GSILRIFDSQ VDSSAPLPTE NKEETLVSAG VQINMSSPTP  
 1151 NKDKAVDTPV LADIISITVD LSSFVPEQDG TLPLPPEIII PKGTXLHNSA  
 1201 IDLKIIDPTN VGYENHALLS SHKDIPLISL KTAEGMTGTP TADASLSNIK  
 1251 IDVSLPSITP ATYGHGTGWS ESKMEDGRIV VGWQPTGYKL NPEKQALVL  
 1301 NNLWSHYTDL RALKQEIFAH HTIAQRMELD FSTNVWGSGL GVVEDCQNIQ  
 1351 EFDGFKHHLT GYALGLDTQL VEDFLIGGCF SQFFGKTESQ SYKAKNDVKS  
 1401 YMGAAYAGIL AGPWLIKGA FVYGNINNDLT TDYGTGLIST GSWIGKGFIA  
 1451 GTSIDYRYIV NPPRFISAIV STVVPFVEAE YVRIDLPEIS BQGKEVRTFQ  
 1501 KTRFENVAIP PGFALEHAYS RGSRAEVNSV QLAVVFDVYR KGPVSLITLK  
 1551 DAAYSWSYSG VDIPCKAWKA RLSNNTTEWNS YLSTYLAIFY EWREDLIAYD  
 1601 FNGGIRIIF\*

A predicted signal peptide is highlighted.

The cp7287 nucleotide sequence <SEQ ID 84> is:

20 1 ATGGTAGCGA AAAAAACAGT ACGATCTTAT AGGCTCTTCAT TTTCTCATTC  
 51 CGTAATAGTA GCAATATTGT CAGCAGGCAT TGCTTTTGAA GCACATTCCCT  
 101 TACACAGCTC AGAAGTAGAT TTAGGTGTAT TCAATAAACA GTTTGAGGAA  
 151 CATTCTGCTC ATGTTGAAGA GGCTCAAACA TCTGTTTAA AGGGATCAGA  
 201 TCCTGTAAAT CCCTCTCAGA AAGAATCCGA GAAGGTTTGT TACACTCAAG  
 251 TGCTCTCTAC CCAAGGAAGC TCTGGAGAGA GTTTGATCT CGCCGATGCT  
 25 301 AATTTCTTAG AGCATTTTCA GCATCTTTT GAAGAGACTA CAGTATTTGG  
 351 TATCGATCAA AAGCTGGTTT GGTCAGATT AGATACTAGG AATTTTCCC  
 401 AACCCACTCA AGAAGCTGAT ACAAGTAATG CTGTAAGTGA GAAATCTCC  
 451 TCAGATACCA AAGAGAATAG AAAAGACCTA GAGACTGAAG ATCCTTCAA  
 501 AAAAAAGTGGC CTTAAGAAG TTTTCATCAGA TCTCCCTAAA AGTCCCTGAAA  
 30 551 CTGCAGTAGC AGCTATTTCT GAAGATCTTG AAATCTCAGA AAACATTTCA  
 601 GCAAGAGATC CTCTTCAGGG TTTAGCATTT TTTTATAAAA ATACATCTTC  
 651 TCAGTCTATC TCTGAAAAGG ATTCTTCATT TCAAGGAATT ATCTTTTCTG  
 701 CTTCAGGAGC TAATTGAGGG CTAGGTTTGT AAAATCTTAA GGCGCCGAAA  
 751 TCTGGGGCTG CAGTTTATTC TGATCGAGAT ATTGTTTGT AAAATCTTGT  
 35 801 TAAAGGATTG AGTTTATAT CTGTGTAATC TTTAGAGAT GGCTCTGCCG  
 851 CAGGTGTAAG CATTGTTGTG ACCCATTTGT GTGATGTAAC TCTCACATGAT  
 901 TGTGCCACTG GTTTAGACCT TGAAGCTTTA CGTCTGTTA AAGATTTTTC  
 951 TCGTGGAGGA GCTGTPTTCA CTGCTCGCAA CCATGAAGTG CAAAATAACC  
 1001 TGCAGGTGG AATCTATACC GTTGTAGGCA ATAAAGGAGC TATTGTTGTA  
 40 1051 GAGAAAAATA GTGCTGAGAA GTCCAAATGGA GGAGCTTTTG CTTGCGGAAG  
 1101 TTTTGTATTAC AGTAACAACG AAAACACCGC CTGTGGAAG GAAATCAAG  
 1151 CATTATCAGG AGGAGCCATA TCCTCAGCAA GTGATATTGA TATTCAAGGG  
 1201 AACTGTAGCG CTATTGAATT TTCAGGAAAC CAGTCTCTAA TTGCTCTTGG  
 1251 AGAGCATATA GGGCTTACAG ATTTTGTAGG TGGAGGAGCT TTAGCTGCTC  
 45 1301 AAGGGACGCT TACCTTAAGA AATAATGCAG TAGTGCAATG TGTAAAAAC  
 1351 ACTTCTAAAA CACATGGTGG AGCTATTTTA GCAGGTACTG TTGATCTCAA  
 1401 CGAAACAATT AGCGAAGTTG CCTTTAAGCA GAATACAGCA GCTCTAACTG  
 1451 GAGGTGCTTT AAGTGCAAAAT GATAAGTTA TAATTGCAAA TAACCTTTGGA  
 1501 GAAATCTTTT TTGAGCAAAA CGAAGTGAGG AATCAGGAG GAGCCATTTA  
 50 1551 TTGTGGATGT CGATCTAATC CTAAGTTAGA ACAAAGGAT TCTGGAGAGA  
 1601 ACATCAATAT TATTGGAAC TCCGGAGCTA TCACTTTTTC AAAAAATAAG  
 1651 GCTTCTGTTT TAGAAGTGAT GACACAAGCT GAAGATTATG CTGGTGAGG  
 1701 CGCTTTATGG GGGCATAATG TTCTTCTAGA TTCCAATAGT GGAATATTC  
 1751 AATTTATAGG AAATATAGGT GGAAGTACCT TCTGGATAGG AGAATATGTC  
 55 1801 GGTGGTGGTG CGATTCTCTC TACTGATAGA GTGACAATTT CTAATACTC  
 1851 TGGAGATGTT GTTTTAAAG GAAACAAAG CCAATGTCTT GCTCAAAAAT  
 1901 ATGTAGCTCC TCAAGAAACA GCTCCCGTGG AATCAGATGC TTCATCTACA  
 1951 AATAAAGACG AGAAGAGCCT TAATGCTTGT AGTCATGGAG ATCATTATCC  
 2001 TCTTAAACT GTAGAAGAGG AAGTGCCACC TTCATTGTTA GAAGAACATC  
 60 2051 CTGTTGTTTC TTCGACAGAT ATTGTTGGTG GTGGGGCCAT TCTAGCTCAA  
 2101 CATATCTTTA TTACAGATAA TACAGGAAAT CTGAGATTCT CTGGGAACCT  
 2151 TGGTGGTGGT GAAGAGTCTT CTAAGTCTCG TGATTTAGCT ATCGTAGGAG  
 2201 GAGGTGCTTT GCTTTCTACT AATGAAGTTA ATGTTTGCAG TAACCAAAAT  
 2251 GTTGTPTTTT CTGATAACGT GACTTCAAAT GGTGTGATT CAGGGGGAGC  
 65 2301 TATTTTAGCT AAAAAAGTAG ATATCTCCGC GAACCACTCG GTTGAATTTG



2351 TCTCTAATGG TTCAGGGAAA TTCGGTGGTG CCGTTTGCGC TTAAACGAA  
 2401 TCAGTAAACA TTACGGACAA TGGCTCGGCA GTATCATTTCT CTAATAATAG  
 2451 AACACGTCCTT GCGGCTGCTG GAGTTGCAGC TCCTCAAGGC TCTGTAACGA  
 2501 TTTGTGGAAA TCAGGGAAAC ATAGCATTTA AAGAGAACTT TGTTTTGGC  
 5 2551 TCTGAAAAATC AAAGATCAGG TGGAGGAGCT ATCATTGCTA ACTCTTCTGT  
 2601 AAATATTTCAG GATAACGCAG GAGATATCCT ATTTGTAAGT AACTCTACGG  
 2651 GATCTTATGG AGGTGCTATT TTTGTAGGAT CTTTGGTTGC TTCGTAAGGC  
 2701 AGCAACCCAC GAACGCTTAC AATTACAGGC AACAGTGGGG ATATCCTATT  
 10 2751 TGCTAAAAATC AGCACGCAAA CAGCCGCTTC TTTATCAGAA AAAGATTCTT  
 2801 TGGGTGGAGG GGCCATCTAT ACACAAAACC TCAAAATTGT AAAGAATGCA  
 2851 GGGAACGTTT CTTCTATGAG CAACAGAGCT CCTAGTGGTG CTGGTGTCCTA  
 2901 AATTGCAGAC GGAGGAACATG TTTGTTTAGA GGCTTTTGGG GGAGATATCT  
 2951 TATTTGAAGG GAATATCAAT TTTGATGGGA GTTTCAATGC GATTCACCTA  
 3001 TGGGGGAATG ACTCAAAAAAT CGTAGAGCTT TCTGCTGTTT AAGATAAAAA  
 15 3051 TATTTATTTTC CAAGATGCAA TTACTTATGA AGAGAACACA ATTCTGCGCT  
 3101 TGCCAGATAA AGATGTACAG CTTTAAAGTG CCCCTTCATT AATTTTAAAC  
 3151 TCCAAGCCAC AAGATGACAG CGCTCAACAT CATGAAGGGA CGATACGGTT  
 3201 TTCTCGAGGG GTATCTAAAA TTCTCAGAT TGCTGCTATA CAAGAGGGAA  
 3251 CCTTAGCTTT ATCACAATAA GCAGAGCTTT GGTGGCAGG ACTTAAACAG  
 20 3301 GAAACAGGAA GTTCTATCGT ATTGCTGCGG GGATCTATTC TCCGTATTTT  
 3351 TGATTCCCAG GTTGATAGCA GTGCGCTCT TCCTACAGAA AATAAAGAGG  
 3401 AGACTCTTGT TTTGCGCGGA GTTCAAATTA ACATGAGCTC TCCTACACCC  
 3451 AATAAAGATA AAGCTGTAGA TACTCCAGTA CTTGCAGATA TCATAAGTAT  
 3501 TACTGTAGAT TTGTTCTCAT TTGTTCTTGA GCAAGACGGA ACTCTTCCTC  
 25 3551 TTCTCTCTGA AATTATCATT CCTAAGGGAA CAAAATTACA TTCTAATGCC  
 3601 ATAGATCTTA AGATTATAGA TCCTACCAAT GTGGGATATG AAAATCATGC  
 3651 TCTTCTAAGT TCTCATAAAG ATATTCCATT AATTTCTCTT AAGACAGCGG  
 3701 AAGGAATGAC AGGGACGCTT ACAGCAGATG CTTCTCTATC TAATATAAAA  
 3751 ATAGATGTAT CTTTACCTTC GATCACACCA GCAACGTATG GTCACACAGG  
 30 3801 AGTTTGGTCT GAAAGTAAAA TGGAAGATGG AAGACTTGTG GTCGGTTGGC  
 3851 AACCTACGGG ATATAAGTTA AATCTGAGA AGCAAGGGGC TCTAGTTTTG  
 3901 AATAATCTCT GGAGTCATTA TACAGATCTT AGAGCTCTTA AGCAGGAGAT  
 3951 CTTTGCTCAT CATACGATAG CTCAAAGAAT GGAGTTAGAT TTCTCGACAA  
 4001 ATGCTCGGGG ATCAGGATTA GGTGTGTGTG AAGATTGTCA GAACATCGGA  
 35 4051 GAGTTTGTAT GGTTCAAACA TCATCTACA GGGTATGCC TAGGCTTGGG  
 4101 TACACAAGTA GTTGAAGACT TCTTAATTGG AGGATGTTTC TCACAGTTCT  
 4151 TTGGTAAAC TGAAAGCCAA TCCTACAAAG CTAAGAACGA TGTGAAGAGT  
 4201 TATATGGGAG CTGCTTATGC GGGGATTTTA GCAGGTCTCT GGTAAATAAA  
 4251 AGGAGCTTTT GTTTACGGTA ATATAAACAA CGATTTGACT ACAGATTACG  
 40 4301 GTACTTTAGG TATTTCAACA GGTTCATGGA TAGGAAAAGG GTTTATCGCA  
 4351 GGCACAAGCA TTGATFACCG CTATATTGTA AATCCTCGAC GGTTTATATC  
 4401 GGCAATCGTA TCCACAGTGG TTCTTTTGT AGAAGCCGAG TATGTCCGTA  
 4451 TAGATCTTCC AGAAATFAGC GAACAGGGTA AAGAGGTTAG AACGTTCCAA  
 4501 AAAACTCGTT TTTGAGAATGT CGCCATTCTT TTTGGATTGG CTTTGAACA  
 45 4551 TGCTTATTCG CGTGGCTCAC GTGCTGAAGT GAACAGTGTG CAGCTTGCTT  
 4601 ACGTCTTTGA TGTATATCGT AAGGGACCTG TCTCTTTGAT TACACTCAAG  
 4651 GATGCTGCTT ATTCFTGGAA GAGTTATGGG GTAGATATTC CTTGTAAAGC  
 4701 TTGGAAGGCT CGCTTGAGCA ATAATACGGA ATGGAATTCA TATTTAAGTA  
 4751 CGTATTTAGC GTTTAATTAT GAATGGAGAG AAGATCTGAT AGCTTATGAC  
 50 4801 TTCAATGGTG GTATCCGTAT TATTTTCTAG

The PSORT algorithm predicts an inner membrane location (0.106).

The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 42A. The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 42B) and for FACS analysis (Figure 42C). A his-tagged protein was also expressed.

55 The cp7287 protein was also identified in the 2D-PAGE experiment and showed good cross-reactivity with human sera, including sera from patients with pneumonitis.

These experiments show that cp7287 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

**Example 43**

The following *C.pneumoniae* protein (PID 4377105) was expressed <SEQ ID 85; cp7105>:

```

1  MSLYQXWVNS QLKSLCYST VAALIFMIPS QESFADSLID LNLGLDPSVE
51  CLSGDGAFSV GYFTKAGSTP VEYQPFKYDV SKRTFTILSV ETANQSGYAY
101 GISYDGTITV GTCSLGAGKY NGAKWSADGT LTPLTGITGG TSHTEARAIS
151 KDTQVIEGFS YDASGQPKAV QWASGATTVT QLADISGGSR SSYAYALSDD
201 GTIIVGSMES TITRKTIAVK WNNVPTYLK TLGGDASTGL YISGDGTVIV
251 GAANTATVTN GNQESHAYMY KDNQMKD*

```

The cp7105 nucleotide sequence <SEQ ID 86> is:

```

10 1  GTGAGTCTAT ATCAAAAATG GTGGAACAGT CAGTTAAAGA AGAGCCTCTG
51  CTATTCGACT GTTGCTGCTC TAATATTTAT GATTCTTCTT CAAGAATCCT
101 TTGCAGATAG TCTTATAGAT TTAAATTTAG GTTTAGATCC TTCGGTCGAA
151 TGTCTGTCTAG GAGATGGTGC ATTTTCTGTT GGGTATTTTA CTAAGGCGGG
201 ATCGACTCCC GTAGAATATC AGCCGTTTAA ATACGACGTA TCTAAGAAGA
15 251 CATTACAAT CCTTCCGTA GAAACGGCAA ATCAGAGCGG CTATGCTTAC
301 GGAATCTCCT ACGATGGCAC GATCACTGTA GGAACGTGTA GCCTAGGTGC
351 AGGAAAATAT AACGGCGCAA AATGGAGTGC GGATGGCACT TTAACACCCT
401 TAACTGGAAT CACGGGGGGG ACGTCACATA CGGAAGCGCG TCGATTCTT
451 AAGGATACTC AGGTGATCGA GGGTTTCTCA TATGATGCTT CAGGGCAACC
20 501 CAAGGCTGTG CAGTGGGCAA GCGGAGCGAC TACAGTAACA CAATTAGCAG
551 ATATTTTCAGG AGGCTCTAGA AGCTCTTATG CGTATGCTAT ATCTGATGAT
601 GGCACGATTA TTGTTGGGTC TATGGAGAGC ACGATAACAA GGAAACTAC
651 AGCTGTAAAA TGGGTAAATA ATGTTCTTAC GTATCTGGGA ACCTTAGGAG
701 GAGATGCTTC TACAGGTCTT TATATTTCTG GAGACGGCAC CGTGATTGTA
25 751 GGTGCGGCAA ATACAGCAAC TGTAACCAAT GGAATCAGG AATCCACGCG
801 CTATATGTAT AAAGATAACC AAATGAAAGA TTGA

```

The PSORT algorithm predicts an inner membrane location (0.100).

The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 43A.

The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 43B) and for FACS analysis (Figure 43C). A his-tagged protein was also expressed.

This protein also showed good cross-reactivity with human sera, including sera from patients with pneumonitis.

These experiments show that cp7105 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

**Example 44**

The following *C.pneumoniae* protein (PID 4376802) was expressed <SEQ ID 87; cp6802>:

```

1  MSNQLQPCIS LGCVSYINSF PLSLQLIKRN DIRCVLAPPA DLLNLLIEGK
51  LDVALTSSLG AISHNLGYVP GFGIAANQRI LSVNLYAAPT FFNSPQPRIA
101 ATLESRSSIG LLKVLCRHLW RIPTPHILRF ITTKVLRQTP ENYDGLLLIG
40 151 DAALQHPVLP GFVTYDIASG WYDLTKLPFV FALLLHSTSW KEHPLPNLAM
201 EEALQQFESS PEEVLKEAHQ HTGLPPSLIQ EYYALCQYRL GEEHYESFEK
251 FREYYGTLYQ QARL*

```

A predicted signal peptide is highlighted.

The cp6802 nucleotide sequence <SEQ ID 88> is:

```

45 1  ATGTCTAACC AACTCCAGCC ATGTATAAGC TTAGGCTGCG TAAGTTATAT
51  TAATTCCTTT CCGCTGTCCC TACAATCAT AAAAAAGAAC GATATTGCT
101 GTGTTCTTGC TCCCCGTGCA GACCTCCTCA ACTTGCTAAT CGAAGGAAAA
151 CTCGATGTTG CTTTGACCTC ATCCCTAGGA GCTATCTCTC ATAACCTGGG
201 GTATGTCCCC GCCTTTGGAA TTGCAGCAAA CCAACGTATC CTCAGTGTA

```

-84-

251 ACCTCTATGC AGCTCCCACT TTCTTTAACT CACCGCAACC TCGGATTGCC  
 301 GCAACTTTAG AAAGTCGCTC CTCTATAGGA CTCTTAAAAG TGCTTTGTCTG  
 351 TCATCTCTGG CGCATCCCAA CTCCTCATAT CCTAAGATTC ATAAC TACAA  
 401 AAGTACTCAG ACAAACCCCT GAAATTTATG ATGGCCTCCT CCTAATCGGA  
 451 GATGCAGCGC TACAACATCC TGTACTTCCT GGATTGTAA CCTATGACCT  
 501 TGCCTCGGGG TGGTATGATC TTACAAAGCT ACCTTTTGTA TTGCTCTTC  
 551 TTCTACACAG CACCTCTTGG AAAGAACATC CCCTACCCAA CCTTGCATG  
 601 GAAGAAGCCC TCCAACAGTT CGAATCTTCA CCCGAAGAAG TCCTTAAAGA  
 651 AGCTCATCAA CATAACAGTC TGCCCCCTTC TCTTCTTCAA GAATACTATG  
 701 CCCTATGCCA GTACCGTCTA GGAGAAGAAC ACTACGAAAG CTTTGAAAAA  
 751 TTCCGGGAAT ATTATGGAAC CCTCTACCAA CAAGCCGAC TGTA

The PSORT algorithm predicts an inner membrane location (0.060).

The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 44A.

The recombinant protein was used to immunise mice, whose sera were used in a Western blot

15. (Figure 44B) and for FACS analysis (Figure 44C). A his-tagged protein was also expressed.

These experiments show that cp6802 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

#### Example 45

The following *C.pneumoniae* protein (PID 4376390) was expressed <SEQ ID 89; cp6390>:

20 1 MVFSYYCMGL FFFSGAISSC GLLVSLGVGL GLSVLGVLLL LLAGLLLFKI  
 51 QSMLEREVPKA PDLLDLEDAS ERLRVKASRS LASLPKEISQ LESYIRSAAN  
 101 DLNTIKTTPH KDQRLVETVS RKLERLAAQ NYMISELCEI SEILREEHH  
 151 LILAQRSLEW IGKSLFSTFL DMESFLNLSH LSEVRPYLAV NDPRLLEITE  
 201 ESWEVVSHPF NVTSAFKKAQ ILFRKNEHSR MKKKLRSVQE LLETFLYKSL  
 25 251 KRSYRELGCL SEKMRIHDN PLFPWVQDQ KYAHARNEFG EICARCLEEFE  
 301 KTFFWLDEEC AISYMDCWDF LNESIQNKKS RVDRDYISTK KIALKDRART  
 351 YAKVLLLENP TTEGKIDLQD AQRAPERQSQ EFTYLTETET KVRLEALQQC  
 401 FSDLREATNV RQVRFTNSEN ANDLKESEFK IDKERVRYQK EQLRYWETID  
 451 RNEQELREBI GESLRLQNR KGYRAGYDAG RLKGLLRQWK KNLRDVEAHL  
 30 501 EDATMDFEHE VSKSELCSVR ARLEVLEEEEL MMSPKVADI KELLSEYERC  
 551 ILPIRENLER AYLQYNKCE ILSKARFFFP EDEQLLVSEA NLREVGQOLK  
 601 QVQGKQERA QKFAIFEKHI QEQKSLIKEQ VRSFDLAGVG FLKSELLSIA  
 651 CNLYIKAVVK ESIPVDVPCM QLYYSYEDN EAVVRNRLN MTERYQNFKR  
 701 SLNSIQFNGD VLLRDFVYQP BGHETRLKER ELQETLSCK KLKVAQDRLS  
 35 751 KLESRLSRR

A predicted signal peptide is highlighted.

The cp6390 nucleotide sequence <SEQ ID 90> is:

40 1 TTGGTATTCT CATACTATTG CATGGGATTA TTTTTTTTCT CTGGAGCTAT  
 51 TTCTAGTTGT GGTCTTTTAG TGTCTCTAGG AGTTGGTTTA GGACTTAGTG  
 101 TTTTAGGAGT ACTTTTACTT CTCCTTAGCAG GTCTTTTGCT TTTTAAGATC  
 151 CAAAGTATGC TTCGAGAGGT GCCTAAGGCT CCTGATCTAT TAGATTTAGA  
 201 AGATGCAAGT GAACGGCTTA GAGTAAAGGC TAGCCGTCTT TTAGCAAGCC  
 251 TCCCGAAGGA AATCAGTCAG CTAGAGAGCT ACATTCGTTC TGCAGCTAAT  
 301 GATCTTAAATA CAATTAAGAC TTGGCCGCAT AAAGATCAAA GACTCGTCGA  
 45 351 GACCGTGTCA CGAAAATTAG AGCGTCTGGC AGCTGCTCAA AACTATATGA  
 401 TTTCTGAAC CTGCGAGATT AGTGAGATTC TTGAGGAAGA GGAGCATCAT  
 451 CTAATTTTGG CTCAGGAATC TCTAGAATGG ATAGGTAAGA GTCTATTTTC  
 501 TACCTTTCTG GACATGGAAT CTTTTTTAAA TTTGAGCCAT CTATCTGAAG  
 551 TGCGTCCGTA CTTAGCTGTA AATGATCCTA GATTATTAGA AATTACCGAA  
 50 601 GAATCTTGGG AAGTAGTGAG TCATTTTATA AATGTAACGT CTGCTTTTAA  
 651 GAAAGCTCAG ATTCTTTTAA AGAACAACGA ACATTCCTCG ATGAAGAAGA  
 701 AGTTAGAAAG TGTTCAGAG TTAGTGAAA CATTTATTTA TAAGAGTTTA  
 751 AAGAGAAGTT ATCGAGAATT AGGATGCTTA AGTGAAAAGA TGAGAATCAT  
 801 TCACGACAAT CCTCTCTTCC CTTGGGTGCA AGATCAGCAG AAGTATGCTC  
 55 851 ATGCTAAGAA TGAATTTGGA GAGATTGCGC GGTGTTTAGA GGAGTTTGAA  
 901 AAGACGTTCT TCTGGTTGGA TGAGGAGTGT GCTATTTCTT ACATGGACTG

-85-

5 951 TTGGGATTTT CTAAATGAGT CTATTCAGAA TAAGAAGTCC AGAGTAGATC  
 1001 GAGATTATAT ATCCACGAAG AAAATTGCAT TAAAGGATAG AGCCCGCACT  
 1051 TATGCTAAGG TTCTTTTAGA AGAGAATCCG ACTACAGAGG GTAAAAATAGA  
 1101 TTTGCAAGAC GCTCAAAGAG CCTTTGAGCG TCAAAGTCAG GAGTTTATA  
 1151 CACTAGAGCA TACGGAAACA AAGGTGAGAC TAGAAGCACT TCAACAGTGC  
 1201 TTCTCGGATC TTAGGGAGGC GACGAACGTA AGGCAAGTTA GGTTTACAAA  
 1251 TTCTGAAAAT GCGAATGATT TAAAGGAGAG TTTTCGAGAAG ATAGATAAAG  
 1301 AGCGTGTGCG ATATCAAAAA GAGCAAAGGC TCTATTGGGA AACAAATAGAT  
 1351 CGCAATGAGC AAGAGCTTAG GGAAGAGATT GGGGAGTCGC TTCGTTTACA  
 1401 AAATCGGAGA AAAGGGTATA GGGCTGGATA TGATGCTGGG CGTTTAAAAG  
 1451 GTTTGTTGCG TCAGTGAAG AAAAATCTCC GCGATGTGGA AGCCACCTT  
 1501 GAAGATGCAA CTATGGATT TGAAGCAAGA GCGAATGTG  
 1551 CAGTGTTCGG GCGAGGCTCG AGGTCTAGA AGAAGAGCTG ATGGATATGT  
 1601 CTCTTAAAGT TCGGGATATA GAAGAGTTGT TGTCTTATGA AGAGCGTTGT  
 1651 ATCTTCTCTA TTAGGAAAAA TTTAGAAAGG GCATACCTCC AATATAATAA  
 1701 GTGTCTCGAA ATTTTATCCA AGGCAAAGTT CTCTTTCCG GAAGACGAGC  
 1751 AATTGCTAGT TTCGGAAGCG AATCTAAGAG AGGTGGGTGC CCAGTTAAAA  
 1801 CAAGTACAGG GAAAATGTCA AGAGAGGGCC CAAAAGTTCC CAATATTGA  
 1851 AAAGCATATT CAGGAGCAGA AAAGCCTTAT TAAAGAGCAA GTCCGGAGTT  
 1901 TTGATCTAGC GGGAGTGGG TTTTAAAGA GTGAGCTTCT TAGTATTGCT  
 1951 TGTAACCTTT ATATAAAGG GGTGTTAAG GAGTCTATAC CAGTTGATGT  
 2001 GCCTTGTATG CAGTTATATT ATAGTTATTA CGAAGATAAT GAAGCTGTAG  
 2051 TCGCAAAACCG CCTTTTAAAT ATGACGGAGA GGTATCAAAA TTTTAAAGG  
 2101 AGTTTGAATT CCATACAATT TAATGGTGAC GTTCTTTTAC GGGATCCGGT  
 2151 CTATCAACCT GAAGGTCATG AGACCAGGCT AAAGGAACGG GAGCTACAAG  
 2201 AAACAACCTT GTCTTGTAAG AAATTTAAAG TGGCTCAAGA TCGTCTTTCT  
 2251 GAATTAGAGT CAAGGCTGTC TAGGAGATAG

The PSORT algorithm predicts a periplasmic location (0.932).

The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 45A.

30 The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 45B) and for FACS analysis (Figure 45C). A his-tagged protein was also expressed.

This protein also showed good cross-reactivity with human sera, including sera from patients with pneumonitis.

35 These experiments show that cp6390 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

#### Example 46

The following *C.pneumoniae* protein (PID 4376272) was expressed <SEQ ID 91; cp6272>:

40 1 MRRCFLFLAS FVIMGSSADA LTHQEAIVKK NSYLSHFPSV SGIVTIEDGV  
 51 LNIHNNLRIO ANKVYVENTV GQSLKLVAHG NVMVNYRAKT LVCDYLEYYE  
 101 DTDSCLLTNG RFAMYPWFLG GSMITLTPET IVIRKGYIST SEGPKKDLCL  
 151 SGDYLEYSSD SLLSIGKTTL RVCRIPIFLP PPFSIMPMEI PKPPINFRGG  
 201 TGGFLGSYLG MSYSPISRKH FSSTFFLDSE FKHGVGMGFN LHCSQKQVPE  
 251 NVFNMKSYIA HRLAIDMAEA HDYRLHGDF CPTHKHVNFV GEYHLSDSWE  
 301 TVADIFPNPF MLKNTGPTRV DCTWNDNYFE GYLTSVVKVN SFQANQELP  
 45 351 YLTLRQYPIS IYNTGVYLEN IVECGYLNFA PSDHIVGENF SSLRLAARPK  
 401 LHKTVPPLPIG TLSSTLGSSL IYSDVPEIS SRHSQLSAKL QLDYRFLHLK  
 451 SYIQRRHIE PFVTFITETR PLAKNEDHYI FSIQDAFHSN NLLKAGIDTS  
 501 VLSKTNPRFP RIHAKLWTH ILSNTESKPT FPKTACELSL PFGKKNVSL  
 551 DAEWIWKHC WDHMNIRWEW IGNDNVAMTL ESLHRSKYSI IKCDRENFIL  
 50 601 DVSRPIDQLL DSPLSDHRNL ILGKLFVRPH PCWNYRLSLR YGWHRQDTPN  
 651 YLEYQMILGT KIFEHWQLYG VYERREADSR FFFFLKLDKP KKPPF\*

A predicted signal peptide is highlighted.

The cp6272 nucleotide sequence <SEQ ID 92> is:

1 ATGAAACGTT GCTTCTTATT TCTAGCTTCC TTGTTCTTA TGGGTTCCCTC

-86-

51 AGCTGATGCT TTGACTCATC AAGAGGCTGT GAAAAAGAAA AACTCCTATC  
 101 TTAGTCACTT TAAGAGTGT TCTGGGATTG TGACCATCGA AGATGGGGTA  
 151 TTGAATATCC ATAACAACCT GCGGATACAA GCCAATAAAG TGTATGTAGA  
 201 AAATACTGTG GGTCAAAGCC TGAAGCTTGT CGCACATGGC AATGTTATGG  
 5 TGAACATATAG GGCAAAAACC CTAGTTTGTG ATTACCTAGA GTATTACGAA  
 301 GATACAGACT CTTGTCTTCT TACTAATGGA AGATTGCGGA TGTATCCTTG  
 351 GTTTCTAGGG GGGTCTATGA TCACTCTAAC CCCAGAAACC ATAGTCATTC  
 401 GGAAGGGATA TATCTCTACC TCCGAGGGTC CCAAAAAGA CCTGTGCCCTC  
 451 TCCGGAGATT ACCTGGAATA TTCTTCAGAT AGTCTTCTTT CTATAGGGAA  
 10 GACAACATTA AGGGTGTGTC GCATTCCGAT ACTTTTCTTA CCTCCATTTT  
 551 CTATCATGCC TATGGAGATC CCTAAGCCTC CGATAAACTT TCGAGGAGGA  
 601 ACAGGAGGAT TTCTGGGATC CTATTTTGGG ATGAGCTACT CGCCGATTTT  
 651 TAGGAAGCAT TTCTCCTCGA CATTTTTCTT GGATAGCTTT TTCAAGCATG  
 701 GCGTCGGCAT GGGATTCAAC CTCCATTGTT CTCAGAAGCA GGTTCCTGAG  
 15 AATGTCCTCA ATATGAAAAG CTATTATGCC CACCGCTTG CTATCGATAT  
 801 GGCAGAAGCT CATGATCGCT ATCGCCTACA CGGAGATTTT TGCTTCACGC  
 851 ATAAGCATGT AAATTTTCTT GGAGAATACC ATCTCAGCGA TAGTTGGGAA  
 901 ACTGTTGCTG ACATTTTCCC CAACAACCTC ATGTTGAAAA ATACAGGCC  
 951 CACACGTGTC GATTGCACTT GGAATGACAA CTATTTTGAA GGGTATCTCA  
 20 CCTCTTCTGT TAAGGTAAAC TCTTTCCAAA ATGCCAACCA AGAGCTCCCT  
 1001 TATTTAACAT TAAGGCAGTA CCCGATTCTT ATTTATAATA CGGGAGTGTA  
 1051 CCTTGAAAAC ATCGTAGAAT GTGGGTATTT AAACTTTGCT TTTAGCGATC  
 1101 ATATCGTTGG CGAGAATTTC TCTTCACTAC GTCTTGCTGC GCGCCCTAAG  
 1151 CTCCATAAAA CTGTGCCTCT ACCTATAGGA ACGCTCTCCT CCACCCTAGG  
 25 GAGTTCCTCTG ATTTACTATA GCGATGTTCC TGAGATCTCC TCGCGCCATA  
 1201 GTCAGCTTTC CGCGAAGCTA CAACTTGATT ATCGCTTCTT ATTACATAAG  
 1251 TCCTACATTC AAAGACGCCA TATTATAGAG CCGTTCGTTA CCTTCATTAC  
 1301 AGAGACTCGT CCTCTAGCTA AGAATGAAGA TCATTATATC TTTTCTATTC  
 1351 AAGATGCCTT TCACTCCTTA AACCTTCTGA AAGCGGGTAT AGATACCTCG  
 1401 GTACTGAGTA AGACTAACCC TCGATTCCCG AGAATCCATG CGAAGCTGTG  
 1451 GACTACCCAC ATCTTGAGCA ATACAGAAAG CAAACCCACG TTTCCCAAAA  
 1501 CTGCACTGCGA GCTATCTCTA CCTTTTGGA AGAAAAATA AGTCTCCTTA  
 1551 GATGCTGAAT GGATTTGGA AAAGCACTGT TGGGATCACA TGAACATACG  
 1601 TTGGGAGTGG ATCGGAAATG ACAATGTGGC TATGACTCTA GAATCCCTGC  
 1651 ATAGAAGCAA ATACAGCCTG ATTAAGTGTG ACAGGGAGAA CTTTCAATTTA  
 1701 ATAGTTCAGC GTCCCATTTGA CCAGCTTTTA GACTCCCTTC TCTCTGATCA  
 1751 TAGGAATCTC ATTTTAGGGA AATTATTTGT ACGACCTCAT CCCTGTGTGA  
 1801 ATTACCGCTT ATCCTTACGC TATGGCTGGC ATCGCCAGGA CACTCCGAAC  
 1851 TACCTAGAAT ACCAGATGAT TCTAGGGACG AAGATCTTCG AACATTGGCA  
 1901 GCTCTATGGG GTGTATGAAC GCCGAGAAGC AGATAGTCGA TTTTCTTCT  
 1951 TCTTAAAGCT CGACAAACCT AAAAAACCTC CCTTCTAA

The PSORT algorithm predicts an outer membrane location (0.48).

The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 46A.

The recombinant protein was used to immunise mice, whose sera were used in a Western blot and for  
 45 FACS analysis (Figure 46B). A his-tagged protein was also expressed.

This protein also showed good cross-reactivity with human sera, including sera from patients with  
 pneumonitis.

These experiments show that cp6272 is a surface-exposed and immunoaccessible protein, and that it  
 is a useful immunogen. These properties are not evident from the sequence alone.

## 50 Example 47

The following *C.pneumoniae* protein (PID 4377111) was expressed <SEQ ID 93; cp7111>:

1 MFEAVIADIQ AREILDSRGY PTLHVKVTTT TGSVGEARVP SGASTGKKEA  
 51 LEFRDFTDPR YQKGVLQAV KNVKEILFPL VKGCSVYEQS LTDSLMMDS  
 101 GSPNKETLGA NAILGVSLAT AHAAAATLRR PLRYRLGGCF ACSLPCPMN  
 55 151 LINGGMHADN GLEFQEFMIR PIGASSIKEA VNMGADVFT LKLLHERGL  
 201 STGVGDEGGF APNLASNEBA LELLLLAIEK AGFTPGKDIS LALDCAASSF

```

251 YNVKTGTYDG RHYEQIAIL SNLCDRYPID SIEDGLAED YDGWALLTEV
301 LGEKVQIVGD DLFVTNPCLI LEGISNGLAN SVLIKPNQIG TLTETVYAIK
351 LAQMAGYTTI ISHRSGETTD TTIADLAVAF NAGQIKTGSL SRSERVAKYN
401 RLMEIEEELG SEAIPTDSNV FSYEDSEE*

```

5 A predicted signal peptide is highlighted.

The cp7111 nucleotide sequence <SEQ ID 94> is:

```

1 ATGTTTGAAG CTGTCATTGC CGATATCCAG GCTAGGGAAA TCTTGGATTG
51 TCGCGGGTAT CCCACTTTAC ATGTTAAAGT AACCACTAGC ACAGGTTCTG
101 TTGGAGAAGC TCGGGTTCCT TCAGGAGCAT CCACAGGGAA AAAAGAAGCC
151 TTAGAGTTTC GTGATACAGA TTCTCCTCGT TATCAAGGCA AAGGGGTTTT
201 GCAAGCTGTA AAAACGTAA AAGAAATCT TTTTCCCCTC GTCAAGGGAT
251 GTAGTGTFTA TGAGCAATCC TTAATTGATT CTCTGATGAT GGATTCTGAC
301 GGCTCTCCGA ACAAGAAAC TCTAGGGGCC AATGCTATTT TAGGAGTCTC
351 TCTAGCTACA GCACATGCAG CAGCAGCAAC ACTACGCAGA CCTCTGTATC
15 401 GTTATTTAGG AGGGTGTTTT GCCTGCAGTC TTCCCTGTCC TATGATGAAT
451 CTGATCAATG GAGGCATGCA TGCCGATAAC GGCTTGGAGT TCCAAGAATT
501 TATGATCCGT CCTATGGAG CCTCTTCCAT CAAAGAAGCT GTCAACATGG
551 GTGCTGACGT TTTTCATACT TTGAAAAAT TACTCCATGA AAGAGGCTTA
601 TCTACTGGAG TGGGTGACGA AGGAGGCTTC GCCCGAATC TTGCTTCTAA
20 651 TGAAGAAGCT CTAGAGCTCC TATTGCTGGC TATTGAAAAA GCAGGCTTTA
701 CTCCAGGAAA AGATATATCG CTAGCCTTAG ACTGCGCAGC ATCCTCATTC
751 TATAACGTAA AAACAGGCAC GTATGATGGG AGGCATATG AAGAGCAAAT
801 CGCAATCCTT TCTAATTTAT GTGATCGCTA TCCTATAGAC TCCATAGAAG
851 ATGGTCTTGC TGAAGAAGAC TATGACGGGT GGGCCTTGT TAACTGAAGTT
25 901 CTTGGAGAAA AAGTACAGAT TGTGGGTGAT GACCTATTG TTACAAATCC
951 GGAATTAATA TTAGAGGGTA TTAGCAATGG ATTAGCGAAC TCTGTGTTGA
1001 TTAACCAAAA TCAGATAGGG ACGCTTACTG AAACAGTGTA TGCTATCAAG
1051 CTTGCGCAAA TGGCTGGCTA TACTACAATT ATTTCTCATC GCTCAGGAGA
1101 AACTACGGAC ACTACGATTG CAGATCTTGC TGTTCCTTTC AACGCCGGTC
30 1151 AAATCAAAAC AGGCTCTTTA TCACGTTCTG AGCGTGTTCG AAAATACAAT
1201 AGACTCATGG AAATGAAGA AGAGCTTGA TCCGAAGCAA TTTTCACAGA
1251 TTCTAATGTA TTTTCTTAC GAGGATTCT GAGGAATAG

```

The PSORT algorithm predicts an inner membrane location (0.100).

The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 47A.

35 The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 47B) and for FACS analysis (Figure 47C). A his-tagged protein was also expressed.

The cp7111 protein was also identified in the 2D-PAGE experiment and showed good cross-reactivity with human sera, including sera from patients with pneumonitis.

40 These experiments show that cp7111 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

#### Example 48

The following *C.pneumoniae* protein (PID 4455886) was expressed <SEQ ID 95; cp0010>:

```

1 MKSQPSWLVL SSTLACFTSC STVFAATAEN IGPDSDFDGS TNTGYTFKN
51 TTTGIDYTLT GDITLQNLGD SAALTRGCFS DTTELSFAG KGYSLSFLNI
45 101 KSSAEGAALS VTDDKNLSLT GFSSLTFLAA PSSVITTPSG KGAVKCGGDL
151 TFDNNGTILF KDYCEENG AISTKNLSLK NSTGSISFEG NKSSATGKKG
201 GAICATGTVD ITNNTAPTLF SNNIAEAAGG AINSTGNCTI TGNTSLVFSE
251 NSVTATAGNG GALSGDADVT ISGNQSVTFS GNQAVANGGA IYAKKLTLAS
301 GGGGVSPFLT IIVQGTAGN GGAISILAAG ECLSLSAEAGD ITFNGNAIVA
50 351 TTPQTTRKNS IDIGSTAKIT NLRAISGHSI FFYDPITANT AADSTDITLNL
401 NKADAGNSTD YSGSVTFSGE KLSDEAKVA DNLSTLTKQP VTLTAGNLVL
451 KRGVTLDTKG FTQTAGSSVI MDAGTTLKAS TEEVTLTGLS IPVDSLGEKG
501 KVVIAASAAS KNVALSGPIL LLDNQGNAYE NHDLGRTQDF SFVQLSALGT

```

551 ATTTDVPAPV TVATPTHYGY QGTWGMTWVD DTASTPKTKT ATLAWNTNGY  
 601 LPNPERQGPL VPNSLWGSFS DIQAIQGVIE RSALTLCSDR GFWAAGVANF  
 651 LDKDKKGEKR KYRHKSGGYA IGGAAQTCSE NLISFAPCQL PGSDKDFLVA  
 701 KNHTDTYAGA FYIQHITECS GFICLLDLKL PGWSHKLPLV LEGQLAYSHV  
 751 SNDLKTQYTA YPEVKGSWGN NAFNMMLGAS SHSYPEYLHC FDTYAPYIKL  
 801 NLTYIRQDSF SEKGTEGRSF DDSNLFNLSL PIGVKFEKFS DCNDFS YDLT  
 851 LSYVPLIRN DPKCTTALVI SGASWETIAN NLARQALQVR AGSHYAFSPM  
 901 FEVLGQFVFE VRGSSRIYNV DLGGKFQF\*

A predicted signal peptide is highlighted.

10 The cp0010 nucleotide sequence <SEQ ID 96> is:

1 ATGAAATCGC AATTTTCCTG GTTAGTGCTC TCTTCGACAT TGGCATGTTT  
 51 TACTAGTTGT TCCACTGTTT TTGCTGCAAC TGCTGAAAAT ATAGGCCCCCT  
 101 CTGATAGCTT TGACGGAAGT ACTAACACAG GCACCTATAC TCCTAAAAAT  
 151 ACGACTACTG GAATAGACTA TACTCTGACA GGAGATATAA CTCTGCAAAA  
 201 CCTTGGGGAT TCGGCAGCTT TAACGAAGGG TTGTTTTCCT GACACTACGG  
 251 AATCTTTAAG CTTTGCCGGT AAGGGGTACT CACTTTCTTT TTTAAATATT  
 301 AAGTCTAGTG CTGAAGGCGC AGCACTTTCT GTTACAACCTG ATAAAAATCT  
 351 GTCGCTAACA GGATTTTCGA GTCTTACTTT CTTAGCGGCC CCATCATCGG  
 401 TAATCACAAC CCCCTCAGGA AAAGGTGCAG TTAATGTGG AGGGGATCTT  
 451 ACATTTGATA ACAATGGAAC TATTTTATTT AAACAAGATT ACTGTGAGGA  
 501 AAATGGCGGA GCCATTTCTA CCAAGAATCT TTCTTTGAAA AACAGCACGG  
 551 GATCGATTTC TTTTGAAGGG AATAAATCGA GCGCAACAGG GAAAAAAGGT  
 601 GGGGCTATTT GTGCTACTGG TACTGTAGAT ATTACAAATA ATACGGCTCC  
 651 TACCCCTCTT TCGAACAATA TTGCTGAAGC TGCAGGTGGA GCTATAAATA  
 25 701 GCACAGGAAA CTGTACAATT ACAGGGAATA CGTCTCTTGT ATTTTCTGAA  
 751 AATAGTGTGA CAGCGACCGC AGGAAATGGA GGAGCTCTTT CTGGAGATGC  
 801 CGATGTTACC ATATCTGGGA ATCAGAGTGT AACTTTCTCA GGAAACCAAG  
 851 CTGTAGCTAA TGGCGGAGCC ATTTATGCTA AGAAGCTTAC ACTGGCTTCC  
 901 GGGGGGGGGG GGGTATCTCC TTTTCTAACA ATAAATAGTCC AAGGTACCAC  
 30 951 TGCAGTAAT GGTGAGCCA TTTCTATACT GGCAGCTGGA GAGTGTAGTC  
 1001 TTTACGAGA AGCAGGGGAC ATTACCTTCA ATGGGAATGC CATTGTTGCA  
 1051 ACTACACCAC AAATACAAA AAGAAATTC ATTGACATAG GATCTACTGC  
 1101 AAAGATCAGC AATTTACGTG CAATATCTGG GCATAGCATC TTTTCTACG  
 1151 ATCCGATTAC TGCTAATACG GCTGCGGATT CTACAGATAC TTTAAATCTC  
 35 1201 AATAAGGCTG ATGCAAGTAA TAGTACAGAT TATAGTGGGT CGATTGTTT  
 1251 TTTTGGTGAA AAGCTCTCTG AAGATGAAGC AAAAGTTGCA GACAACCTCA  
 1301 CTCTTACGCT GAAGCAGCCT GTAACCTAA CTGCAGGAAA TTTAGTACTT  
 1351 AAACGTGGTG TCACTCTCGA TACGAAAGGC TTTACTCAGA CCGCGGGTTC  
 1401 CTCTGTTATT ATGGATGCGG GCACAACGTT AAAAGCAAGT ACAGAGGAGG  
 40 1451 TCACTTTAAC AGGTCTTTCC ATTCTGTAG ACTCTTTAGG CGAGGGTAAG  
 1501 AAAGTTGTAA TTGCTGCTTC TGCAGCAAGT AAAAATGTAG CCTTAGTGG  
 1551 TCCGATTCTT CTTTGTGATA ACCAAGGGAA TGCTTATGAA AATCAGGACT  
 1601 TAGGAAAAAC TCAAGACTTT TCATTGTGTC AGCTCTCTGC TCTGGGTACT  
 1651 GCAACAACCTA CAGATGTTCC AGCGGTTCCCT ACAGTAGCAA CTCCTACGCA  
 45 1701 CTATGGGTAT CAAGTACTT GGGGAATGAC TTGGGTGAT GATACCGCAA  
 1751 GCACTCCAAA GACTAAGACA GCGACATTAG CTTGGACCAA TACAGGCTAC  
 1801 CTTCCGAATC CTGAGCGTCA AGGACCTTTA GTTCTAATA GCCTTTGGGG  
 1851 ATCTTTTCA GACATCCAAG CGATTCAAGG TGTCTAGAG AGAAGTGCTT  
 1901 TGACTCTTTG TTTCAGATCGA GGCTTCTGGG CTGCGGGAGT CGCCAATTTT  
 50 1951 TTAGATAAAG ATAAGAAAGG GGAATAACGC AAATACCGTC ATAAATCTGG  
 2001 TGGATATGCT ATCGGAGGTG CAGCGCAAAC TTGTTCTGAA AACTTAATTA  
 2051 GCTTTGCCTT TTGCCAACTC TTTGGTAGCG ATAAAGATTT CTAGTCGCT  
 2101 AAAAATCATA CTGATACCTA TGCAGGAGCC TTCTATATCC AACACATTAC  
 55 2151 AGAATGTAGT GGGTTCATAG GTTGTCTCTT AGATAAAGT CCTGGCTCTT  
 2201 GGAGTCATAA ACCCTCGTTC TTAGAAGGGC AGCTCGCTTA TAGCCACGTC  
 2251 AGTAATGATC TGAAGACAAA GTATATGCGG TATCTGAGG TGAAAGGTTT  
 2301 TTGGGGGAAT AATGCTTTTA ACATGATGTT GGGAGCTTCT TCTCATTTCT  
 2351 ATCTTGAATA CCTGCATTGT TTTGATACCT ATGCTCCATA CATCAAAGT  
 2401 AATCTGACCT ATATACGTCA GGACAGCTTC TCGGAGAAAG GTACAGAAGG  
 60 2451 AAGATCTTTT GATGACAGCA ACCTCTTCAA TTTATCTTTG CCTATAGGGG  
 2501 TGAAGTTTGA GAAGTTCTCT GATTGTAATG ACTTTTCTTA TGATCTGACT  
 2551 TTTATCTTATG TTCCTGATCT TATCCGCAAT GATCCCAAAT GCACTACAGC  
 2601 ACTTGTAAAT AGCGGAGCCT CTTGGGAAAC TTATGCAAT AACTTAGCAC  
 2651 GACAGGCCTT GCAAGTGCCT GCAGGCAGTC ACTACGCCTT CTCTCCTATG  
 65 2701 TTTGAAGTGC TCGGCCAGTT TGCTTTTGAA GTTCGTGGAT CCTCACGGAT

-89-

2751 TTATAATGTA GATCTTGGGG GTAAGTCCA ATTCTAG

The PSORT algorithm predicts an outer membrane location (0.922).

The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 48A.

The recombinant protein was used to immunise mice, whose sera were used in a Western blot

5 (Figure 48B) and for FACS analysis (Figure 48C). A his-tagged protein was also expressed.

The cp0010 protein was also identified in the 2D-PAGE experiment and showed good cross-reactivity with human sera, including sera from patients with pneumonitis.

These experiments show that cp0010 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

## 10 Example 49

The following *C.pneumoniae* protein (PID 4376296) was expressed <SEQ ID 97; cp6296>:

```

1  MEEVSEYLQQ VENQLESCSK RLTKMETFAL GVRLEAKEEI ESIILSDVVN
51  RFEVLCRDIE DMLSRVEEIE RMLRMAELPL LPIKEALTKA FVQHNSCKEK
101 LTKVEPYFKE SPAYLTSEER LQSLNQTLQR AYKESQKVSF LESEVRACRE
151 QLKDQVRQFE TQGVSLIKEE ILFVTSTFRT KFSYHSFRLH VPCMRLYEY
201 YDDIDLERTR ARWMAMSERY RDAFQAFQEM LKEGLVERAQ ALRETEYWLY
251 REERKSKKKH*
```

The cp6296 nucleotide sequence <SEQ ID 98> is:

```

1  ATGGAGGAGG TGTCTGAGTA TCTTCAGCAA GTAGAAAATC AGTTGGAATC
20  CTGTTCCAAG CGATTAACCA AGATGGAAAC TTTTGCCTTA GGTGTGAGGT
51  TGGAAAGCTAA AGAAGAGATA GAGTCTATCA TACTTTCCTGA TGTAGTGAAC
101 CGTTTTGAGG TTTTATGTAG AGATATTGAA GATATGCTAT CTCGAGTCGA
151 GGAGATAGAG CGGATGTTAC GTATGGCGGA GCTTCCTCTA CTTCCCTATAA
201 AAGAAGCGCT TACCAAGGCT TTTGTACAAC ATAACAGCTG TAAAGAGAAG
251 TTAACCAAGG TAGAGCCTTA CTTTAAAGAG AGCCCTGCAT ATCTAAGTAG
301 TGAAGAGCGA TTGCAGAGTT TGAATCAGAC TTTACAACGT GCGTACAAAG
351 AGTCCCAAAA GGTTCAGGTT TTAGAATCGG AAGTGAGAGC CTGTCGAGAG
401 CAGCTTAAAG ATCAAGTAAG ACAGTTTGAA ACTCAAGGAG TGAGCTTGAT
451 AAAAGAAGAG ATTCTCTTTG TGAAGTAGTAC CTTTAGAACT AAATTTAGCT
501 ATCATTCATT TCGATTACAT GTTCCTTGCA TGAGGTGTGA TGAGGAGTAT
551 TATGATGACA TTGATCTAGA GAGAACTCGA GCTCGATGGA TGGCGATGTC
601 TGAGAGGTAT AGAGATGCTT TTCAGGCATT CCAGGAGATG TTGAAGGAAG
651 GCCTAGTTGA AGAAGCTCAG GCTCTTAGAG AAACCGAGTA CTGGTTATAT
701 CGAGAGGAGA GAAAGAGTAA AAAGAAACAT TGA
751
```

35 The PSORT algorithm predicts a cytoplasmic location (0.523).

The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 49A.

The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 49B) and for FACS analysis (Figure 49C). A his-tagged protein was also expressed.

These experiments show that cp6296 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

## Example 50

The following *C.pneumoniae* protein (PID 4376664) was expressed <SEQ ID 99; cp6664>:

```

1  MVLFAQASG RNRVKADAIV LPFWHFKDAK NAASFEEFE PSYLPALENF
51  QGKTGEIELL YSSPKAKEKR IVLLGLGKNE ELTSDVVFQT YATLTRVLRK
101 AKCSTVNIIL PTISELRLSA EEFLVGLSSG ILSLNYDYPR YNKVDRNLET
```

45



-90-

151 PLSKVTVIGI VPKMADAIFR KEAIFEGVY LTRDLVNRNA DEITPKKLAE  
 201 VALNLGKEFP SIDTKVLGKD AIAKEKMGLL LAVSKGSCVD PHFIVVRYQG  
 251 RPKSKDHTVL IGKGVTFDSG GLDLKPGKSM LTMKEDMAGG ATVLGILSAL  
 301 AVLELPIINV GIIIPATENAI DGASYKMGDV YVGMSGLSVE ICSTDABEGR  
 351 ILADAITYAL KYCKPTRIID FATLTGAMVV SLGEEVAGFF SNNDVLAEDL  
 401 LEASAETSEP LWRLPLVKKY DKTLLHSDIAD MKNLGSNRAG AITAALEFLQR  
 451 FLEESSVAWA HLDIAGTAYH EKEEDRYPKY ASGFGVRSIL YYLENSLSK\*

The cp6664 nucleotide sequence <SEQ ID 100> is:

1 GTGGTTTAT TTCATGCTCA AGCCTCTGGG CGTAATCGTG TTAAGGCAGA  
 51 TGCTATAGTC CTGCCCTTTT GGCATTTTAA GGATGCAAAA AATGCAGCTT  
 101 CTTTGAAGC CGAGTTTGAA CCCTCGTATC TCCCCGCTTT AGAAAACTTT  
 151 CAAGGAAAAA CCGGGGAGAT TGAACCTCCT TATAGTAGTC CTAAAGCTAA  
 201 GGAAAAACGC ATTGTCCTCT TAGGCTTAGG GAAAAATGAA GAGCTCACCT  
 251 CTGATGTTGT TTTCCAAACC TATGCGACAC TAACTCGTGT CTTACGTAAA  
 15 301 GCAAAGTGTT CCACAGTCAA TATCATCTTA CCTACAATTT CTGAATTGCG  
 351 GCTTTCGTCC GAAGAATTCT TAGTGGGGTT GTCCTCAGGA ATTTTCTCAT  
 401 TAAACTATGA CTACCCACGT TATAATAAGG TAGATCGTAA TCTTGAAACT  
 451 CCTCTTTCTA AAGTCACGGT TATCGGTATC GTTCCCAAAA TGGCGGATGC  
 501 TATCTTTAGG AAAGAAGCAG CCATTTTCGA AGGCGTATAT CTCACTCGAG  
 20 551 ATCTTGTAAG CAGGAATGCT GATGAAATTA CCCCTAAGAA ATTGGCAGAG  
 601 GTTGCTCTGA ATCTGGGAAA AGAGTTCCTT AGTATTGATA CTAAGGTCTT  
 651 GGGAAAAGAT GCCATCGCCA AAGAGAAAAT GGGACTCCTA TTGGCTGTTT  
 701 CCAAGGGTTC TTGTGTGGAT CCACACTTTA TCGTTGTCCG TTATCAAGGA  
 751 CGTCTTAAGT CTAAAGATCA CACCGTCTTG ATAGGGAAG GGGTCACTTT  
 25 801 TGACTCTGGA GGTTTAGACC TCAAGCCTGG AAAATCCATG CTTACTATGA  
 851 AAGAAGACAT GGCAGGTGGG GCTACAGTCC TCGGGATTCT CTCGGCGTTA  
 901 GCAGTTTAG AGCTTCCTAT AAATGTCACG GGGATCATTC CTGCTACAGA  
 951 GAATGCTATC GATGGCGCCT CCTATAAAAT GGGAGATGTC TATGTAGGAA  
 1001 TGTGCGGGCT TTCTGTTGAG ATTTGTAGTA CCGATGCTGA GGGACGTCTT  
 30 1051 ATCCTCGCTG ATGCGATTAC ATATGCTTTA AAATATTGTA AACCAGACAG  
 1101 TATTATAGAT TTGCAACTC TAACAGGAGC TATGGTAGTC TCTCTAGGAG  
 1151 AAGAGGTTGC AGGTTTCTTT TCCAATAACG ATGTTTTCAG TGAAGATCTT  
 1201 TTAGAGCGCT CAGCCGAAAC CTCCGAGCCG TTATGGAGAC TTCTCTAGT  
 1251 TAAGAAGTAT GATAAAACAT TGCATTCTGA TATTGCTGAT ATGAAAAATC  
 35 1301 TAGGCAAGTAA CCGTCAGGG GCTATTACAG CAGCATTATT CTTGCAGAGA  
 1351 TTTTGGGAAG AATCTTCGGT AGCTTGGGCA CATCTTGATA TTGCAGGTAC  
 1401 TGCATATCAT GAAAAAGAAG AAGACCGTTA TCCAAAATAT GCTTCAGGTT  
 1451 TTGGTGTCG TTCTATTCTT TATTACTTAG AAAATAGTCT TTCTAAGTAG

The PSORT algorithm predicts an inner membrane location (0.268).

40 The protein was expressed in *E.coli* and purified as a GST-fusion (Figure 50A), as a his-tagged protein, and as a GST/His fusion. The proteins were used to immunise mice, whose sera were used in Western blot Western blot (50B) and FACS (50C) analyses.

The cp6664 protein was also identified in the 2D-PAGE experiment (Cpn0385) and showed good cross-reactivity with human sera, including sera from patients with pneumonitis.

45 These experiments show that cp6664 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 51

The following *C.pneumoniae* protein (PID 4376696) was expressed <SEQ ID 101; cp6696>:

50 1 MTLIFVIIIV WCNAPLIKLC VIMGLQSRLO HCIEVSQNSN FDSQVKQFIY  
 51 ACQDKFTLRQS VLKIFRYHPL LKIHDIARAV YLLMALEEGB DLGLSFLNVQ  
 101 QYPSGAVELF SCGGFPWKGL PYPAEHAEFG LLLLQIAEFY RESQAYVSKM  
 151 SHFQQALFDH QGSVFPSSLWS QENSRLLEKE TTLSQSFLFQ LGMQIHPEYS  
 201 LEDPALGFWM QTRSSSAFV AASGCQSSLG AYSSGDVGVI AYGPCSGDIS  
 251 DCYFYGCCGI AKEFVCQKSH QTTEISFLTS TGKPHPRNTG PSYLKDSYVH  
 55 301 LPIRCKITIS DKQYRVHAAL AEATSAMTFS IFCKGKNCQV VDGPRLRSCS

351 LDSYKPGND IMILGENDAI NIVSASPYME IFALQGKEKF WNADFLINIP  
401 YKEEGVMLIF EKKVTSEKGR FFTKMN\*

A predicted signal peptide is highlighted.

The cp6696 nucleotide sequence <SEQ ID 102> is:

```

5      1  TTGACTCTAA  TTTTGTGTAT  TATTATCGTT  TGGTGCAATG  CTTTTCTGAT
      51  CAAATGTGTC  GTGATAATGG  GGCTGCAATC  CAGGTTACAA  CATTGTATAG
     101  AAGTGTCCCA  GAATTCGAAC  TTTGATTAC   AAGTAAAACA  GTTTATCTAT
     151  GCGTGCCAAG  ATAAGACATT  AAGGCAGTCT  GTACTCAAGA  TTTTCCGCTA
     201  CCATCCTTTA  CTA AAAAATTC  ATGATATTGC  TCGGGCCGTC  TATCTTTTGA
    10      251  TGGCCTTAGA  AGAAGGCGAG  GATTTAGGCT  TAAGCTTTTT  AAATGTACAG
     301  CAGTACCCCT  CAGGTGCTGT  AGAACTGTTT  TCTTGTGGGG  GATTTCTTTG
     351  GAAAGGATTA  CCTTATCCTG  CAGAACATGC  GGAATTTGGC  CTACTCCTGT
     401  TACAGATCGC  AGAGTTTAT  GAAGAGAGTC  AGGCATACGT  CTCTAAAATG
     451  AGTCATTTTC  AACAGGCACT  CTTTGATCAC  CAAGGGAGCG  TCTTTCCCTC
    15      501  TCTCTGGAGC  CAGGAGAACT  CTCGACTCCT  AAAAGAAAAG  ACAACTCTTA
     551  GCCAATCGTT  TCTCTTCCAA  TTAGGAATGC  AAATTCACCC  AGAATACAGT
     601  CTTGAGGATC  CTGCACTAGG  GTTCTGGATG  CAAAGAACGC  GTTCTTCATC
     651  CGCTTTTGTA  GCCGCTTCAG  GATGTCAAAG  TAGCTTGGGA  GCGTATTCCT
     701  CAGGGGATGT  CGGTGTTATC  GCTTATGGAC  CTTGCTCTGG  AGACATTAGT
    20      751  GATTGTATT  ATTTTGGATG  TTGTGGAATC  GCTAAAGAGT  TCGTGTGCCA
     801  AAAATCTCAC  CAAACTACAG  AGATTCTCTT  TCTCACCTCT  ACAGGAAAGC
     851  CTCATCCCAG  AAATACGGGA  TTTTCTTACC  TTCGAGATTC  CTATGTACAT
     901  CTGCCGATCC  GCTGTAAGAT  CACTATTTCC  GACAAGCAAT  ATCGCGTGCA
     951  CGCTGCGTTG  GCTGAGGCCA  CCTCTGCCAT  GACGTTTCT  ATTTTCTGTA
    25     1001  AGGGGAAGAA  TTGTCAGGTT  GTTGACGGCC  CTCGCTTGCG  CTCCTGTTCC
    1051  CTAGATTCTT  ATAAAGGTCC  CGGAAACGAC  ATTATGATTC  TTGGGGAAAA
    1101  TGACGCAATC  AACATTGTTT  CTGCAAGTCC  CTATATGGAA  ATTTTGTGCT
    1151  TGCAAGGCAA  AGAAAAATTT  TGGAATGCAG  ACTTTTGAT  TAATATTCCT
    1201  TACAAAGAAG  AGGGCGTCAT  GTTAATTTT  GAAAAAAAAG  TGACCTCTGA
    30     1251  GAAAGGAAGA  TTC TTTACGA  AGATGAATTA  A

```

The PSORT algorithm predicts an inner membrane location (0.463).

The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 51A. The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 51B) and for FACS analysis (Figure 51C). A his-tagged protein was also expressed.

35 This protein also showed good cross-reactivity with human sera, including sera from patients with pneumonitis.

These experiments show that cp6696 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

#### Example 52

40 The following *C.pneumoniae* protein (PID 4376790) was expressed <SEQ ID 103; cp6790>:

```

      1  MSEHRKSSKI  IGIDLGTINS  CVSVMEGGQA  KVITSSEGTR  TTPSIVAFKG
     51  NEKLVGIPAK  RQAVTNPEKT  LGSTKRFIGR  KYSEVASEIQ  TVPYTVTSGS
    101  KGDVAFVDG  KQYTPBEIGA  QILMKMKETA  BAYLGETVTE  AVITVPAYFN
    151  DSQRASTKDA  GRIAGLDVGR  IPEPTAAAL  AYGIDKVGDK  KIAVFDLGGG
    201  TFDISILEIG  DGVFEVLSTN  GDTLLGGDDF  DEVIIKWMIE  EFKKQEGIDL
    251  SKDNMALQRL  KDAAEKAKIE  LSGVSSTEIN  QPFITMDAQQ  PKHLALTLTR
    301  AQFERLAASL  IERTKSPCIK  ALSDAKLSAK  DDDVLLVGG  MSRMPAVQET
    351  AKELFGKEPN  KGVNPDEVVA  IGAAIQGGVL  GGEVKDVLLE  DVIPLSLGIE
    401  TLGGVMTTLV  ERNTTIPTQK  KQIFSTAADN  QPAVTIVVLQ  GERPMAKDNK
    451  EIGRFDLTDI  PPAPRGHPQI  EVSFDIDANG  IFHVSADIVA  SGKEQKIRIE
    501  ASSGLQDEI  QRMVRDAEIN  KEEDKKRREA  SDAKNEADSM  IFRAEKAIKD
    551  YKEQIPETLV  KEIEERIENV  RNALKDDAPI  EKIKEVTEDL  SKHMOKIGES
    601  MQSQSASAAA  SSAANAKGGP  NINTEDLKKH  SFSTKPPSNN  GSSEDHIEEA

```

651 DVEIIDNDDK\*

The cp6790 nucleotide sequence &lt;SEQ ID 104&gt; is:

```

      1  ATGAGTGAAC  ACAAAAAATC  AAGCAAATTT  ATAGGTATAG  ACTTAGGCAC
5      51  AACAAACTCC  TGCCTATCTG  TTATGGAAGG  AGGACAAGCT  AAAGTAATTA
      101  CATCATCCGA  AGGAACAAGA  ACCACGCCAT  CGATCGTTGC  CTTCAAAGGT
      151  AATGAGAAAT  TAGTGGGGAT  TCCAGCAAAA  CGTCAAGCAG  TGACAAATCC
      201  AGAAAAAACT  CTCGGCTCTA  CAAAACGCTT  TATTGGCCGT  AAGTACTCTG
      251  AAGTAGCTTC  GGAATCCAA  ACCGTTCCTT  ATACAGTCAC  CTCCGGATCT
      301  AAAGGTGATG  CCGTTTTCGA  AGTTGATGGC  AAACAATACA  CTCCAGAAGA
10     351  AATTGGCGCA  CAAATCTTAA  TGAAAATGAA  AGAGACAGCA  GAAGCTTATC
      401  TAGGCGAAAC  TGTACAGAA  GCAGTGATCA  CCGTCCCCGC  ATACTCAAT
      451  GATTCTCAAC  GAGCATCCAC  AAAAGATGCT  GGACGCATTG  CAGGTCTAGA
      501  TGTA AACCGT  ATCATTCAG  AACCTACCGC  AGCAGCTCTT  GCCTACGGAA
      551  TCGATAAAGT  CCGTGATAAA  AAAATCGCTG  TCTTCGACCT  TGGTGGAGGA
15     601  ACTTTTGATA  TCTCCATCCT  AGAAATCGGT  GATGGCGTCT  TCGAAGTTCT
      651  ATCTACAAAT  GGAGATACTC  TCCTCGGTGG  AGACGACTTT  GATGAAGTCA
      701  TTATCAAATG  GATGATCGAA  GAATTCAAAA  AACAAGAAGG  CATTGATCTT
      751  AGCAAAGATA  ATATGGCCTT  ACAAAGACTT  AAAGATGCTG  CTGAGAAAGC
      801  AAAAAATGAA  CTTTCAGGAG  TCTCTTCCAC  AGAAATCAAT  CAGCCATTCA
20     851  TCACAATGGA  TGCACAAGGA  CCTAAACACC  TTGCATTGAC  ACTCACACGT
      901  GCGCAATTCG  AGAAACTCGC  AGCCTCTCTA  ATCGAAAGAA  CAAAATCTCC
      951  ATGCATCAAA  GCACTCAGTG  ACGCAAAACT  TTCCGCTAAG  GATATCGATG
100    1001  ATGTCTCTTT  AGTTGGAGGT  ATGTCAAGAA  TGCCCGCAGT  GCAAGAAACT
      1051  GTAAAAGAAC  TCTTCGGCAA  AGAGCCTAAT  AAAGGAGTCA  ACCCCGACGA
25     1101  AGTTGTTGCT  ATTGGAGCCG  CAATTCAAGG  TGGTGTCTTT  GGCGGAGAAG
      1151  TTAAGGATGT  TCTACTTCTA  GACGTTATCC  CCCTATCTCT  GGGTATCGAA
      1201  ACTCTAGGAG  GCGTCATGAC  GACTCTGGTA  GAGAGAAATA  CTACAAATCCC
      1251  TACACAGAAA  AAACAAATCT  TCTCCACAGC  TGCTGATAAC  CAGCCTGCGG
      1301  TTACCATCGT  AGTTCTCCAA  GGAGAGCGTC  CCATGGCCAA  AGATAACAAG
30     1351  GAAATCGGAA  GATTGATCT  TACAGATATC  CCTCCGGCTC  CTCGAGGCCA
      1401  TCCTCAAATC  GAAGTCTCCT  TCGATATCGA  TGCAAAACGA  ATTTTCCATG
      1451  TCTCAGCTAA  AGATGTTGCC  AGCGGTAAAG  AACAGAAAAT  TCGTATCGAA
      1501  GCAAGCTCAG  GACTTCAAGA  AGATGAAATC  CAAAGAATGG  TTCGAGATGC
35     1551  CGAAATTAAT  AAGGAAGAAG  ATAAAAACG  TCGTGAAGCT  TCAGATGCTA
      1601  AAAATGAAGC  CGATAGCATG  ATCTTCAGAG  CCGAAAAAGC  TATTAAAGAT
      1651  TATAAGGAGC  AAATTCCTGA  AACTTTAGTT  AAAGAAATCG  AAGAGCGAAT
      1701  CGAAAACGTG  CGCAACGCAC  TCAAAGATGA  CGCTCCTATT  GAAAAAATTA
      1751  AAGAGGTTAC  TGAAGACCTA  AGCAAGCATA  TGCAAAAAAT  TGGAGAGTCT
40     1801  ATGCAATCGC  AGTCTGCATC  AGCAGCAGCA  TCATCGGCAG  CCAATGCTAA
      1851  AGGTGGACCT  AACATCAATA  CAGAAGATT  T  GAAAAACAT  AGTTTCAGTA
      1901  CGAAGCCTCC  TTCAAATAAC  GGTTCTTCAG  AAGACCATAT  CGAAGAAGCT
      1951  GATGTAGAAA  TTATTGATAA  CGACGATAAG  TAA

```

The PSORT algorithm predicts an inner membrane location (0.151).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 52A) and a his-tagged product. The proteins were used to immunise mice, whose sera were used in Western blot (Figure 52B) and FACS (Figure 52C) analyses.

The cp6790 protein was also identified in the 2D-PAGE experiment (Cpn0503).

These experiments show that cp6790 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

## 50 Example 53

The following *C.pneumoniae* protein (PID 4376878) was expressed <SEQ ID 105; cp6878>:

```

      1  MNVPDSKNLH  PPAYELLEIK  ARITQSYKEA  SAILTAIPDG  ILLLSETGHF
51  LICNSQAREI  LGIDENLEIL  NRSFTDVLPD  TCLGFSIQEA  LESLKVPKTL
101  RLSLCKESKE  KEVELFIRKN  EISGYLFIQI  RDRSDYKQLE  NAIRYKNIA
55  151  ELGKMTATLA  HEIRNPLSGI  VGFASILKKE  ISSPRHQRL  SSIISGTRSL
      201  NNLVSSMLEY  TKSQPLNLKI  INLQDFFSSL  IPLLSVSFPN  CKFVREGAQP

```

-93-

251 LFRSIDPDRM NSVVWNLVKN AVETGNSPIT LTLHTSGDIS VTNPGTIPSE  
 301 IMDKLFPPFF TTKREGNGLG LABAQKIIRL HGGDIQLKTS DSAVSFFIII  
 351 PELLAALPKE RAAS\*

The cp6878 nucleotide sequence <SEQ ID 106> is:

5 1 ATGAACGTCC CTGATTCCAA GAACCTCCAT CCTCCTGCAT ACGAACTCCT  
 51 AGAGATCAAG GCTCGCATCA CACAATCTTA TAAAGAAGCG AGTGCTATAC  
 101 TGACAGCGAT TCCTGATGGT ATCCTATTAC TTTCTGAAAC AGGACACTTT  
 151 CTTATCTGCA ATTCAACAAGC ACGTGAAATT CTAGGAATTG ATGAAAATCT  
 201 AGAAATCTTT AATAGATCCT TTACCGATGT TCTCCCCGAT ACGTGTCTTG  
 10 251 GATTTCCTAT TCAAGAGGCT CTTGAATCTC TAAAGTCCC TAAACTCTTT  
 301 AGACTCTCTC TCTGTAAAGA ATCTAAAGAA AAAGAAGTGG AACTCTTCAT  
 351 CCGTAAAAAC GAGATCAGTG GATACCTGTT TATCCAAATC CGCGATCGGT  
 401 CCGACTATAA ACAACTAGAA AACGCTATAG AAAGATATAA AAATATCGCA  
 451 GAACCTGGGA AAATGACGGC TACCCTAGCT CACGAAATCC GCAATCCGCT  
 15 501 AAGTGGAAAT GTTGGATTTG CCTCTATCCT AAAGAAAGAG ATTTCCTCTC  
 551 CTCGCCACCA ACGAATGCTC TCCTCAATCA TCTCCGGCAC AAGGTCTCTA  
 601 AATAACCTTG TCTCTCTAT GTTAGAATAT ACAAAATCAC AACCGTGA  
 651 CCTAAAGATT ATAAATTTAC AAGACTTCTT CTCTTCTCTT ATCCCTCTGC  
 701 TCTCCGTCTC TTTCCCGAAT TGCAAGTTTG TAAGAGAGGG CGCACAACT  
 20 751 CTATTCAGAT CTATAGATCC TGATCGGATG AACAGTGTG TTTGGAACCT  
 801 AGTGAAAAAT GCTGTAGAAA CAGGGAAGTC TCCGATCACT CTGACCCTGC  
 851 ATACATCGGG AGACATCTCG GTAACGAACC CCGGAACGAT TCCTTCCGAG  
 901 ATCATGGACA AGCTCTTCAC TCCATTCTTC ACAACAAAGA GAGAGGGAAA  
 951 TGGTTTGGGA CTTGCTGAAG CTCAAAAAAT TATAAGACTC CATGGAGGAG  
 25 1001 ATATCCAATT AAAACAAGC GACTCCGCCG TTAGCTTCTT CATAATCATC  
 1051 CCCGAACTTC TAGCGGCCCT ACCCAAAGAA AGAGCCGCTA G

The PSORT algorithm predicts an inner membrane location (0.204).

The protein was expressed in *E.coli* and purified as a his-tag product (Figure 53A) and as a GST-fusion product. The recombinant GST-fusion protein was used to immunise mice, whose sera were used in a Western blot (Figure 53B) and for FACS analysis.

These experiments show that cp6878 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

#### Example 54

The following *C.pneumoniae* protein (PID 4377224) was expressed <SEQ ID 107; cp7224>:

35 1 MMKIKRKVAL AVGGSGGHIV PALSVKEAFS REGIDVILLG KGLKNHPSLQ  
 51 QGISYREIPS GLPTVLNPIK IMSRTLSLCS GYLKARKELK IFDPLVIGF  
 101 GSYHSLFVLL AGLSHKIPLF LHEQNLVPGK VNQLFSRYAR GIGVNFSPVT  
 151 KHFRCPAEV FLPKRSFSLG SPMKRCRTH TPTICVVGGS QGAQILNFCV  
 201 PQALVKLVNK YPNMYVHHIV GPKSDVMKVQ HVYNRGEVLC CVKPFEEQLL  
 40 251 DVLLAADLVI SRAGATILEE ILWAKVPGIL IPYPGAYGHQ EVNAKFFVDV  
 301 LEGGTMILEK ELTEKLLVEK VTFALDSHNR EKQRNSLAAY SQQRSTKTFH  
 351 AFICECL\*

The cp7224 nucleotide sequence <SEQ ID 108> is:

45 1 ATGATGAAGA AAATTCGAAA AGTAGCCTTG GCTGTAGGAG GTTCAGGAGG  
 51 CCACATGTGC CCAGCTCTCT CCGTAAAGGA AGCTTTTCTT CGTGAAGGAA  
 101 TAGACGTATT ACTACTAGGG AAAGGTCTCA AGAACCATCC TTCTTTGCAA  
 151 CAGGGAATCA GCTATCGGGA AATCCCTTCA GGAATCCTA CAGTCCTTAA  
 201 TCCCATAAAG ATCATGAGCA GGACCCCTTC TCTATGTTCA GGATACCTGA  
 251 AAGCAAGAAA GGAACCTAAA ATTTTGGACC CTGACCTGGT CATAGGATTT  
 50 301 GGGAGCTACC ACTCTCTTCC CGTGTGCTC GCAGGACTGT CCCATAAAAT  
 351 TCCCTTATTT CTACACGAAC AAAATCTAGT TCCTGGAAAA GTAAATCAAT  
 401 TGTTTTCCCG CTATGCTCGA GGTATTGGAG TGAATTTCTC CCCCCTTACT  
 451 AAGCACTTCC GCTGCCCCGC AGAAGAGGTC TTCCTTCTTA AACGAAGCTT  
 501 CTCCTTAGGA AGCCCTATGA TGAAGCGATG TACAAATCAT ACCCTACAA  
 55 551 TCTGTGTTGT TGGAGGTTCT CAGGGAGCAC AGATATTAAA TACTTGTGTT  
 601 CCCCAAGCTC TTGTCAAGCT AGTCAATAAG TACCCAAATA TGTACGTCCA

-94-

```

651 TCATATTGTA GGACCTAAAA GTGATGTTAT GAAGGTGCAA CATGTTTACA
701 ATCGTGGAGA GGTCTCTGCG TGTGTGAAGC CGTTTGAAGA GCAACTCCTA
751 GATGTCCTTC TTGCCGCGAGA TTTGGTTCATC AGTAGGGCAG GAGCCACAAT
801 TTTAGAAGAA ATTCTTTGGG CAAAAGTTCC CGGAATTTTA ATTCCCTATC
851 CAGGAGCTTA TGGACATCAG GAAGTTAATG CTAAATTCCT TGTAGACGTC
901 TTAGAAGGGG GAACTATGAT CCTAGAAAAA GAATTAACAG AGAAGCTATT
951 AGTAGAAAAA GTAACGTTTG CTTTAGACTC CCATAACAGA GAAAAACAAC
1001 GCAATTCCCT AGCGGCGTAT AGTCAGCAA GGTCAACAAA AACATTCCTT
1051 GCATTCATTT GTGAATGCTT ATAG

```

10 The PSORT algorithm predicts an inner membrane location (0.164).

The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 54A. The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 54B) and for FACS analysis (Figure 54C). A his-tagged protein was also expressed.

15 This protein also showed good cross-reactivity with human sera, including sera from patients with pneumonitis.

These experiments show that cp7224 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 55

The following *C.pneumoniae* protein (PID 4377140) was expressed <SEQ ID 109; cp7140>:

```

20 1 MVRRSISFCL FFLMTLLCCT SCNSRSLIVH GLPGREANEI VVLLVSKGVA
51 AOKLPOAAAA TAGAATEQMW DIAVPSAQIT EALAILNQAG LPRMKGTSLL
101 DLFAKQGLVP SELQEKIRYQ EGLSEQMAST IRKMDGVVDA SVQISFTTEN
151 EDNLPLTASV YIKHRGVLDN PNSIMVSKIK RLIASAVPGL VPENVSUVSD
201 RAAYSDITIN GPWGLTEIID YVSVWGIILA KSSLTKFRLI FYVLILILFV
25 251 ISCGLLWVIW KTHTLIMTMG GTRGFFNPTP YTKNALEAKK AEGAAADKEK
301 KEDADSQGES KNAETSDKDS SDKDAPEGSN EIEGA*

```

A predicted signal peptide is highlighted.

The cp7140 nucleotide sequence <SEQ ID 110> is:

```

30 1 ATGGTTCGTC GATCTATTTT TTTTGTCTTG TCTTTCTTAA TGACATGCT
51 GTGTGTGACA AGCTGTAAAC GCAGGTCTCT AATTGTGCAC GGTCTTCCTG
101 GCAGAGAAGC GAATGAGATT GTGGTGCTTT TGGTAAGCAA AGGGGTGGCT
151 GCACAAAAAT TGCCTCAAGC TGCAGCGGCT ACAGCCGGAG CAGCTACTGA
201 GCAAATGTGG GATATCGCGG TTCCGTCAGC ACAAATCACA GAGGCCCTTG
25 251 CCATTCTAAA TCAAGCGGGT CTTCCACGTA TGAAAGGGAC AAGCCTGTTA
301 GATCTTTTTG CAAAACAAGG TCTTGTTCCT TCCGAGCTTC AGGAAAAAAT
35 351 CCGTTATCAA GAAGGCTTAT CAGAACAGAT GGCTCTACG ATTAGAAAAA
401 TGGATGGCGT TGTGATGCC TCAGTACAGA TTTCTTTCAC TACAGAAAAA
451 GAAGATAATC TTCTTTTAAC AGCCTCTGTG TATATTAAGC ATCGAGGGGT
501 TTTGGACAAT CCGAACAGCA TTATGGTTTC CAAAATTAAG CGCCTTATTG
40 551 CAAGTGCTGT TCCAGGACTT GTGCCAGAGA ACGTCTCTGT AGTGAGCGAT
601 CGCGCAGCTT ATAGTGATAT TACAATTAAT GGTCTTTGGG GATTACAGA
651 AGAAATCGAT TATGTTTCTG TTTGGGGTAT TATTCTTGGC AAGTCTTCGC
701 TCACCAAAAT CCGTCTCATT TTTTATGTCT TGATTCTCAT TTTATTGTGT
751 ATTTCTTGTC GTCTCCTTTG GGTCAATTGG AAAATCATA CTCTCATTAT
45 801 GACTATGGGA GGTACAAAAG GGTCTTTCAA CCCTACACCA TATACAAAAG
851 ATGCCTTGGA AGCCAAGAAA GCCGAGGGAG CAGCTGCTGA CAAAGAGAAA
901 AAAGAAGATG CAGATTACCA GGGGGAAGC AAAATGCGG AAACCAGTGA
951 TAAAGACTCT AGTGATAAAG ATGCTCCAGA AGGAAGCAAT GAAATTGAGG
1001 GTGCTTAG

```

50 The PSORT algorithm predicts an inner membrane location (0.650).

The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 55A. The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 55B) and for FACS analysis (Figure 55C). A his-tagged protein was also expressed.

These experiments show that cp7140 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

#### Example 56

The following *C.pneumoniae* protein (PID 4377306) was expressed <SEQ ID 111; cp7306>:

```

1  MITKQLRSLW AVLVGSSLLA LPLSGQAVGK KESRVSELPQ DVLLKEISGG
51  FSKVATKATP AVVYIESFPK SQAVTHPSPG RRGPYENPFD YFNDEFFNRF
101 FGLPSQREKP QSKEAVRGTG FLVSPDGYIV TNNHVVEDTG KIHVTLHDGQ
151 KYPATVIGLD PKTDLAVIKI KSONLPYLSF GNSDHLKVG D WAIAIGNPFG
201 LQATVTVGVI SAKGRNQLHI ADFEDFIQTD AAINPGNSGG PLLNIDGQVI
251 GVNTAIVSGS GGYIGIGFAI PSLMANRIID QLIRDGQVTR GFLGVTLQPI
301 DAELAACYKL EKVGALVTD VVKGSPADKA GLKQEDVILIA YNGKEVDSLS
351 MFRNAVSLMN PDTRIVLKVV REGKVIEIPV TVSQAPKEDG MSALQRVGIR
401 VQNLTPETAK KGLIAPETKG ILIISVEPGS VAASSGLAPG QLILAVNRQK
451 VSSIEDLNRT LKDSNNENIL LMVSQGDVIR FIALKPPE*

```

A predicted signal peptide is highlighted.

The cp7306 nucleotide sequence <SEQ ID 112> is:

```

1  ATGATAACTA AGCAATFGCG TTCGTGGCTA GCTGTACTTG TTGGTTCAAG
51  TCTGCTAGCT CTTCCTTTAT CAGGGCAAGC TGTCGGGAAA AAAGAATCTC
101 GAGTTTCCGA GCTGCCTCAA GACGTTCTTC TTAAAGAGAT CTCGGGAGGG
151 TTTTCTAAGG TCGCTACCAA GGCGACTCCC GCTGTGTGTG ACATAGAAAG
201 TTTCCCAAAG AGCCAGGCTG TAACACATCC TTCTCCTGGA CGCCGTGGGC
251 CTTATGAAAT TCCTTTTGAT TATTTAATG ATGAGTTTTC CAATCGTTTC
301 TTTGGTCTAC CTTACACAGAG GGAAAAACCT CAAAGTAAAG AGGCGGTTCG
351 AGGAACAGGT TTCCTAGTAT CTCCAGATGG CTATATTGTG ACTAATAACC
401 ATGTTGTCTG AGATACAGGT AAGATTCACG TAACTCTTCA TGATGGGCAA
451 AAGTACCCAG CAACTGTAAT CGGACTCGAT CCTAAAACAG ACCTTGCAST
501 CATTAAATTT AAATCCCAAA ACCTCCCGTA TCTTCTTTT GGAAACTCCG
551 ACCACTTAAA AGTCGGAGAT TGGGCAATTG CAATTGGAAA TCCCTTCGGT
601 CTTCAAGCTA CGGTCAACCGT AGGTGTCATC AGTGCTAAAG GAAGAAATCA
651 ACTCCACATT GCAGATTTTG AAGATTTTAT TCAGACAGAT GCTGCGATTA
701 ATCCAGGCAA CTCTGGAGGC CCTCTTCTAA ATATTGATGG ACAGGTCATC
751 GGTGTTAATA CTGCCATTGT CAGTGGTAGT GGTGGCTATA TTGGAATCGG
801 GTTTGCGATT CCTAGCCTTA TGGCAAATAG AATCATAGAT CAGCTGATTC
851 GTGATGGTCA AGTTACCCGA GGATTCCTAG GAGTGACTTT ACAACCTATA
901 GATGCGGAAC TCGCTGCTTG CTACAACTC GAAAAGGTTT ATGGCGCTTT
951 AGTCACAGAT GTTGTAAAG GATCTCCAGC AGATAAAGCA GGGCTAAAC
1001 AAGAAGATGT GATCATTGCT TATAATGGGA AAGAAGTCGA TTCACTGAGT
1051 ATGTTCCGTA ATGCTGTTTC TTAAATGAAT CCAGATACAC GTATTGTTCT
1101 AAAGGTAGTT CGTGAAGGAA AGGTTATCGA AATACCCGTG ACAGTTTCTC
1151 AAGCTCCAAA AGAAGATGGA ATGTCGGCTT TACAGCGTGT GGAATCCGT
1201 GTGCAAAACC TAACTCCTGA AACTGCTAAG AAGCTGGGAA TTGCTCCAGA
1251 GACTAAAGGC ATTTTGATTA TAAGTGTTGA ACCAGGCTCT GTAGCAGCTT
1301 CTTCAGGAAT TGCTCTGGT CAGCTGATCC TTGCTGTGAA TAGACAAAAA
1351 GTATCTTCGA TTGAAGATCT GAATAGAACG TTAAAGATT CTAACAATGA
1401 GAATATCTT CTTATGTTT CTAAGGAGA TGTATTTCGC TTCATTGCC
1451 TGAAACCTGA AGAATAA

```

The PSORT algorithm predicts a periplasmic location (0.923).

The protein was expressed in *E.coli* and purified as a his-tag product (Figure 56A) and as a GST-fusion product (Figure 56B). The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (Figure 56C) and for FACS (Figure 56D) analyses.

The cp7306 protein was also identified in the 2D-PAGE experiment (Cpn0979) and showed good cross-reactivity with human sera, including sera from patients with pneumonitis.

These experiments show that cp7306 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

## 5 Example 57

The following *C.pneumoniae* protein (PID 4377132) was expressed <SEQ ID 113; cp7132>:

```

1  MCNSTIAMKKQ KRGFVLMELL MSFTLIAALLL GTLGFWYRKI YTVQKQKERI
51  YNFYIEESRA YKQLRTLFSM SLSSSYEEPG SLFSLIFDRG VYRDPKLAGA
101 VRASLHHDTK DQRLRLRICN IKDQSYFETQ RLLSHVTHVV LSFQRNPDPB
151 KLPETIALTI TREPKAYPPR TLTYQFAVGK*

```

A predicted signal peptide is highlighted.

The cp7132 nucleotide sequence <SEQ ID 114> is:

```

1  ATGTGTAACCT CTATAGCTAT GAAAAAGCAA AAGCGTGGCT TTGTGCTTAT
51  GGAAATTACTC ATGTCGTICA CTCTAATTGC TTGTTTATTA GGGACTTTAG
101 GATTTTGGTA TCGGAAAATT TATACTGTAC AAAAGCAAAA AGAACGTATT
151 TATAACTTTT ATATCGAAGA AAGCCGAGCC TACAAGCAGC TCAGAACCCCT
201 GTTTAGCATG TCCTTGCTCT CATCTTACGA GGAGCCTGGA TCATTATTTT
251 CTTTAACTCT TGATCGGGGT GTTTATCGAG ATCCTAAGCT GGCAGGTGCG
301 GTACGAGCTT CTCTCCATCA TGACACCAAG GATCAGAGAT TGGAACTTCG
351 TATTTGTAAT ATTAAGGATC AGTCTTACTT TGAAACACAG CGACTGCTCT
401 CCCACGTGAC CCATGTTGTA CTTTCCTTCC AGAGAAATCC TGATCCTGAA
451 AAACCTCCTG AAACAATGTC TTTAACATA ACACGGGAAC CTAAAGCATA
501 TCCTCCAAGG ACGTTAACAT ACCAATTGCG GGTGGGAAA TAA

```

The PSORT algorithm predicts a periplasmic location (0.915).

25 The protein was expressed in *E.coli* and purified as a his-tag product (Figure 57A) or as a GST-fusion. The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (Figure 57B) and FACS (Figure 57C) analyses.

These experiments show that cp7132 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

## 30 Example 58

The following *C.pneumoniae* protein (PID 4376733) was expressed <SEQ ID 115; cp6733>:

```

1  MKTSIPWVLV SSVLAFSCHL QSLANEELLS PDSFNGNID SGTFTPKTSA
51  TTYSLTGDFV FYEPGKGTPD SDSCFKQTTD NLTFGLNGHS LTFGFI DAGT
101 HAGAAASTTA NKNLTFSGFS LLSFDSSPST TVTTGQGTLS SAGGVNLENI
35 151 RKLVVAGNFS TADGGAIKGA SFLLTGTS GD ALFSNNSSST KGGAIATTAG
201 ARIANNFGYV RFLSNIASTS GGAIDDEGTS ILSNNKFLYF EGNAAKTGG
251 AICNTRKASGS PELIISNNKT LIPASNVAET SGGAIHAKKL ALSSGGFTEF
301 LRNVSSATP KGAISIDAS GELSLSAETG NITFVRNLT TTGSTDTPKR
351 NAINIGSNKG FTELRAAKNH TIFYDPITS EGTSSDV LKI NNGSAGALNP
40 401 YQGTILPSGE TLTADELKVA DNLKSSFTQP VSLSGGKLLL QRGVTLBSTS
451 FSQEAGSLLG MDSGTTLSTT AGSITITNLG INVDSLGLKQ PVSLTAKGAS
501 NKVIVSGKLN LIDIEGNIYE SHMFSDQLF SLLKITVDAD VDTNVDISSL
551 IPVPAEDPNS EYGFQGWNV NWTDTATNT KBATATWTKT GFVPSPERKS
601 ALVCNTLWGV FTDIRSLQQL VEIGATGMEH KQGFVWSSMT NFLHKTGDEN
45 651 RKGFRHTSGG YVIGGSAHTP KDDLFTFAPC HLFARDKDCF IAHNNSRTYG
701 GTLFFPKSHT LQPQNYLRLG RAKFSESAIE KFPREIPLAL DVQVSFSSHSD
751 NRMETHYTSL PESEGSWSNE CIAGGIGLDL PFVLSNPHEL FKTFTIPQMKV
801 EMVYVSQNSF FESSSDGRGF SIGRLNLNLSI PVGAKFVQGD IGSYTYDLS

```

-97-

851 GFFVSDVYRN NPQSTATLVM SPDSWKIRGG NLSRQAFLLR GSNNYVYNSN  
 901 CELFGHYAME LRGSSRNYNV DVGTKLRF\*

A predicted signal peptide is highlighted.

The cp6733 nucleotide sequence <SEQ ID 116> is:

```

5      1  ATGAAGACTT CGATTCCCTTG GGTTTTAGTT TCCTCCGTGT TAGCTTTCTC
      51  ATGTCACCTA CAGTCACTAG CTAACGAGGA ACTTTTATCA CCTGATGATA
     101  GCTTTAATGG AAATATCGAT TCAGGAACGT TTA CTCCAAA AACTTCAGCC
     151  ACAACATATT CTCTAACAGG AGATGTCTTC TTTTACGAGC CTGGAAAAGG
     201  CACTCCCTTA TCTGACAGTT GTTTTAAAGCA AACCACGGAC AATCTTACCT
10    251  TCTTGGGGAA CGGTCAATAGC TTAACGTTTG GCTTTATAGA TGCTGGCACT
     301  CATGCAGGTG CTGCTGCATC TACAACAGCA AATAAGAATC TTACCTTCTC
     351  AGGGTTTTCC TTA CTGAGTTT TGATTCTCTC TCCTAGCACA ACGGTTACTA
     401  CAGGTCAAGG AACGCTTTCC TCAGCAGGAG GCGTAAATTT AGAAAATATT
     451  CGTAAACTTG TAGTTGCTGG GAATTTTCTT ACTGCAGATG GTGGAGCTAT
15    501  CAAAGGAGCG TCTTTCCCTT TAACTGGCAC TTCTGGAGAT GCTCTTTTTA
     551  GTAACAACCT TTCATCAACA AAGGGAGGAG CAATTGCTAC TACAGCAGGC
     601  GCTCGCATAG CAAATAACAC AGGTTATGTT AGATTCTTAT CTAACATAGC
     651  GTCTACGTCA GGAGGCGCTA TCGATGATGA AGGCACGTCG ATACTATCGA
     701  ACAACAAATT TCTATATTTT GAAGGGAATG CAGCGAAAAC TACTGGCGGT
20    751  GCGATCTGCA ACACCAAGGC GAGTGGATCT CCTGAACTGA TAATCTCTAA
     801  CAATAAGACT CTGATCTTTG CTTCAAACGT AGCAGAAACA AGCGGTGGCG
     851  CCATCCATGC TAAAAAGCTA GCCCTTTTCT CTGGAGGCTT TACAGAGTTT
     901  CTACGAAATA ATGTCTCATC AGCAACTCCT AAGGGGGGTG CTATCAGCAT
     951  CGATGCCTCA GGAGAGCTCA GTCTTTCTGC AGAGACAGGA AACATTACCT
25   1001  TTGTAAGAAA TACCCCTTACA ACAACCGGAA GTACCGATAC TCCTAAACGT
     1051  AATGCCATCA ACATAGGAAG TAACGGGAAA TTCACGGAAT TACGGGCTGC
     1101  TAAAAATCAT ACAATTTTCT TCTATGATCC CATCACTTCA GAAGGAACCT
     1151  CATCAGACGT ATTGAAGATA AATAACGGCT CTGCGGGAGC TCTCAATCCA
     1201  TATCAAGGAA CGATTCTATT TTCTGGAGAA ACCCTAACAG CAGATGAAC
30   1251  TAAAGTTGCT GACAAATTTAA AATCTTCATT CACGCAGCCA GTCTCCCTAT
     1301  CCGGAGGAAA GTTATTGCTA CAAAAGGGAG TCACTTTAGA GAGCACGAGC
     1351  TTCTCTCAAG AGGCCGGTTC TCTCCTCGGC ATGGATTTCAG GAACGACATT
     1401  ATCAACTACA GCTGGGAGTA TTACAATCAC GAACCTAGGA ATCAATGTTG
35   1451  ACTCCTTAGG TCTTAAGCAG CCCGTCAGCC TAACAGCAA AAGGTGCTTCA
     1501  AATAAAGTGA TCGTATCTGG GAAGCTCAAC CTGATTGATA TTGAAGGGAA
     1551  CATTATGATA AGTCATATGT TCAGCCATGA CCAGCTCTTC TCTCTATTAA
     1601  AAATCACGGT TGATGCTGAT GTTGATACTA ACGTTGACAT CAGCAGCCTT
     1651  ATCCCTGTTC CTGCTGAGGA TCCTAATTCA GAATACGGAT TCCAAGGACA
     1701  ATGGAATGTT AATTGGAATA CGGATACAGC TACAAATACA AAAGAGGCCA
40   1751  CGGCAACTTG GACCAAAACA GGATTTGTTT CCAGCCCCGA AAGAAAATCT
     1801  GCGTTTAGTAT GCAATACCCCT ATGGGGAGTC TTTACTGACA TTCGCTCTCT
     1851  GCAACAGCTT GTAGAGATCG GCGCAACTGG TATGGAACAC AAACAAGGTT
     1901  TCTGGGTTTC CTCCATGACG AACTTCCTGC ATAAGACTGG AGATGAAAAT
     1951  CGCAAAGGCT TCCGTCATAC CTCTGGAGGC TACGTCTATG GTGGAAGTGC
45   2001  TCACACTCCT AAAGACGACC TATTTACCTT TCGGTTCTGC CATCTCTTTG
     2051  CTAGAGACAA AGATTGTTTT ATCGCTCACA ACAACTCTAG AACCTACGGT
     2101  GGAATTTTAT TCTTCAAGCA CTCTCATACC CTACAACCCC AAAACTATTT
     2151  GAGATTAGGA AGAGCAAAGT TTTCTGAATC AGCTATAGAA AAATTCCCTA
     2201  GGGAAATTCC CCTAGCCTTG GATGTCCAAG TTTCTGTCAG CCATTTCAGC
50   2251  AACCCTATGG AAACGCACTA TACCTCATTG CCAGAATCCG AAGGTTCTTG
     2301  GAGCAACGAG TGTATAGCTG GTGGTATCGG CCTAGACCTT CCTTTTGTTC
     2351  TTTCCAACCC ACATCCTCTT TTCAAGACCT TCATTCCACA GATGAAAGTC
     2401  GAAATGGTTT ATGTATCACA AAATAGCTTC TTCGAAAGCT CTAGTGATGG
     2451  CCGTGGTTTT AGTATTGGAA GGCTGCTTAA CCTCTCGATT CCGTGGGTG
55   2501  CGAAATTCGT GCAGGGGGAT ATCGGAGATT CCTACACCTA TGATCTCTCA
     2551  GGATTCCTTG TTTCCGATGT CTATCGTAAC AATCCCCAAT CTACAGCGAC
     2601  TCTTGTGATG AGCCCAGACT CTGGAATAAT TCGCGGTGGC AATCTTTCAA
     2651  GACAGGCATT TTTACTGAGG GGTAGCAACA ACTACGTCTA CAACTCCAAT
     2701  TGTGAGCTCT TCGGACATTA CGCTATGGAA CTCGCTGGAT CTTCAAGGAA
60   2751  CTACAAATGA GATGTTGGTA CCAAACCTCG ATTCTAG
  
```

The PSORT algorithm predicts an outer membrane location (0.924).



The protein was expressed in *E.coli* and purified as a his-tag product, as shown in Figure 58A. The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 58B) and for FACS (Figure 58C) analyses. A GST-fusion protein was also expressed.

The cp6733 protein was also identified in the 2D-PAGE experiment (Cpn0451).

- 5 These experiments show that cp6733 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 59

The following *C.pneumoniae* protein (PID 4376814) was expressed <SEQ ID 117; cp6814>:

```

10      1  MHDALLSILA IQELDIKMIR LMRVKEHQK ELAKVQSLKS DIRRKVQEK
      51  LEMENLKTQI RDGENRIQEI SEQINKLENQ QAAVKKMDFF NALTQEMTTA
      101  NKERRSLEHQ LSDLMDKQAG GEDLIVSLKE SLASTENSSS VIEKEIFESI
      151  KKINEEGKAL LEQRTKELKHA TNPELLSIYE RLLNNKKDRV VVPIENRVCS
      201  GCHIVLTPQH ENLVRKKDRL IFCKHCSRIL YWQBSQVNAQ ENSTAKRRRR
      251  RAAV*
```

- 15 The cp6814 nucleotide sequence <SEQ ID 118> is:

```

      1  ATGCATGACG CACTTCTAAG CATTFTGGCT ATTCAAGAGC TTGATATTAA
      51  AATGATTGCG CTTATGCGCG TAAAGAAAGA ACATCAGAAA GAATTGGCTA
      101  AAGTCCAATC TTTAAAAAGT GATATTCGTA GAAAAGTTCA GGAAAAAGAA
      151  CTCGAAATGG AGAATTTGAA AACTCAAATT CGAGATGGAG AGAATCGCAT
      201  CCAAGAGATT TCTGAACAAA TCAATAAATT AGAAAATCAG CAAGCTGCTG
      251  TAAAAAAAT GGATGAGTTT AACGCTCTTA CCCAAGAAAT GACTACAGCA
      301  AACAAAGAAC GTCGCTCTTT AGAGCACCAG CTTAGCGATC TCATGGATAA
      351  GCAAGCTGGA GCGGAAGACC TTATTGCTCT TCTAAAAGAA AGCTTAGCTT
      401  CTACAGAAAA TAGTAGCAGT GTCATTGAAA AAGAAATTTT TGAAAGCATC
      451  AAAAAGATTA ATGAAGAAGG CAAAGCTTTG CTTGAACAAC GGACAGAGTT
      501  AAAGCATGCG ACGAATCCCG AACTACTCAG CATCTATGAG CGTCTATTAA
      551  ACAATAAAAA AGATCGCGTT GTTGTTCCTA TTGAAAATCG TGCTTCGAGT
      601  GGTGTGCATA TTGTTCTAAC TCCTCAACAC GAAAATCTTG TAAGAAAGAA
      651  AGACCGACTC ATTTTTTGCG AACATFGCTC TCGAATCTCT TATTGGCAAG
      701  AATCCCAAGT CAATGCTCAG GAAAATTCCA CAGCAAAACG TCGTCGTCGT
      751  CGCGCAGCTG TATAA
```

The PSORT algorithm predicts an inner membrane location (0.070).

- 35 The protein was expressed in *E.coli* and purified as a GST-fusion (Figure 59A) or his-tagged product. The recombinant proteins were used to immunise mice, whose sera were used in Western blot (Figure 59B) and FACS (Figure 59C) analyses.

These experiments show that cp6814 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 60

The following *C.pneumoniae* protein (PID 4376830) was expressed <SEQ ID 119; cp6830>:

```

40      1  MKWLPAVAVF AAVLPALTAF GDPASVEIST SHTGSGDPTS DAALTGFTQS
      51  STETDGTITY IVGDIFFSTF TNIPVPVVT DANDSSSNSS KGGSSSSGAT
      101  SLIRSSNLHS DFDFTKDSVL DLYHLFFPSA SNTLNPALLS SSSSGGSSSS
      151  SSSSSGSAS AVVAADPKGG AAFYSNEANG TLFTTDSGN PGSLTLQNLK
      201  MTGDGAATYS KGPLVFTGLK NLFTGNEISQ KSGGAAYTEG ALTTQAIVEA
      45  251  VTFTGNTSAG QGGAIYVKEA TLFNALDSLK FEKNTSGQAG GGIYTESTLT
      301  ISNITKSIEF ISNKASVPAP APEPTSPAPS SLINSTTIDT STLQTRAASA
      351  TPAVAVPAV TPTPISTQET AGNGGAIYAK QGISISTFKD LTFKSNSASV
```

-99-

401	DATLTVDSSST	IGESGGAIFA	ADSIQIQQCT	GTTLFSGNTA	NKSGGGIYAV
451	GQVTLIEDIAN	LKMTNNTCKG	EGGAIYTKKA	LTINNGAILT	TFSGNTSTDN
501	GGAIFAVGGI	TLSDLVEVRF	SKNKTGNYS	PITKAASNTA	PVVSSSTTAA
551	SPAVPAAAAA	PVTNAAKGGA	LYSTEGTLVS	GITSILSFEN	NECQNQGGGA
5	601	YVTKTFQCS	SHRLQFTSNK	AADEGGGLYC	GDDVTLTNLT
	651	EKHGGGLSLA	SGKSLTMTSL	ESFCLNANTA	KENGGGANVP
	701	PTPNPAPVQ	QPVYGEALVT	GNTATKSGGG	IYTKNAAFSN
	751	SENGGALLT	QKAADKTDCS	FTYITNVNIT	NNTATGNGGG
	801	IDNLTQVSNQ	AKKGGGVYLE	DALILEKVIT	GSVSQNTATE
10	851	QLQALPGSPT	ITDNKVETSL	TTSTNLYGGG	IYSSGAVTLT
	901	NSVINATATQ	DADIQGGGIY	ATTSLSINQC	NTPILFSNNS
	951	QIAGGAIFSA	AVTIENNSQP	IIFLNNSAKS	EATTAATAGN
	1001	NSVTLTNPE	ITPKGNYAET	GGAIGCIDLT	NGSPPRKVS
	1051	NSALNRGGAI	YGETIDISRT	GATFIGNSSK	HDGSAICCST
15	1101	IFENNKVTET	TATTKASINN	LGAALYGNNE	TSQVTLISLSA
	1151	LCTATNKYCS	IAGNVKFTAI	EASAGKAISF	YDAVNVSTKE
	1201	KATSTGTILF	SGELHENKSY	IPQKVTFAHG	NLILGKNAEL
	1251	TTITMGPQSV	LSNHSKEAGG	IAINNVIIDF	SEIVPTKDNA
	1301	STRNADSKDK	IDITGTVTLL	DPNGNLYQNS	YLGEDRDITL
20	1351	VTATNVTLQG	NLGAKKGYLG	TWNLDPNSSG	SKILKWTFTD
	1401	DNHFYINSIW	GAQNSLVTVK	QGILGNMLNN	ARFEDPAFNN
	1451	RKEVSRNSDS	FTYHGRGYTA	AVDAKPRQEF	ILGAAFSQVF
	1501	NYKHKSGSHS	TQASLYAGNI	FYFPAIRSRP	ILFQGVATYG
25	1551	PSIEKKNMAN	WDSIAWLFDL	RFSVDLKEPQ	PHSTARLTFTY
	1601	EKFTELDYDP	RSFSACSYGN	LAIPTGFSVD	GALAWREIIL
	1651	VILRNPKAT	YEVLSTKEKG	NVVNVLPTRN	AARAEVSSQI
	1701	TYTIDASMNT	LVQMANGGIR	FVF*	YLGSWTLYG

A predicted signal peptide is highlighted.

The cp6830 nucleotide sequence <SEQ ID 120> is:

30	1	ATGAAGTGGC	TACCAGCTAC	AGCTGTTTTT	GCTGCCGTAC	TCCCCGCACT
	51	AACAGCCTTC	GGAGATCCCC	CGTCTGTTGA	AATAAGTACC	AGCCATACAG
	101	GATCCGGGGA	TCCTACAAGC	GACGCTGCCT	TAACAGGATT	TACACAAAGT
	151	TCCACAGAAA	CTGACGGTAC	TACCTATACC	ATTGTCGGTG	ATATCACCTT
	201	CTCTACTTTT	ACGAATATTC	CTGTTCCCGT	AGTAACTCCA	GACGCCAACG
35	251	ATAGTTCCAG	CAATAGCTCT	AAAGGAGGAA	GTAGCAGTAG	TGGAGCTACA
	301	TCTCTAATCC	GATCCTCAAA	CCTACACTCC	GATTTTGTAT	TTACAAAAGA
	351	TACGTGTGTA	GACCTCTATC	ACCTTTCTCT	TCCTTCAGCT	TCAAATACTC
	401	TCAATCCTGC	ACTCCTTTCT	TCCAGTAGCA	GCGGTGGATC	CTCGAGCAGC
	451	AGTAGCTCCT	CATCATCTGG	AAGTGCATCT	GCTGTGTTTG	CTGCGGACCC
40	501	AAAAGGAGGC	GCTGCCTTTT	ATAGTAACGA	GGCTAACGGA	ACTTTAACCCT
	551	TCACCTACGA	CTCTGGAAAT	CCCGGCTCCC	TGACTCTTCA	GAATCTTAAA
	601	ATGACCGGAG	ATGGAGCCGC	CATCTACTCG	AAGGGTCCTC	TAGTATTTAC
	651	TGGTTTAAAA	AATCTAACCT	TTACAGGAAA	TGAATCTCAG	AAATCTGGAG
	701	GTGCTGCCTA	TACTGAAGGC	GCACTCACAA	CACAAGCAAT	CGTTGAAGCC
45	751	GTAACCTTTA	CTGGCAACAC	CTCGGCAGGG	CAAGGAGGCG	CTATCTATGT
	801	TAAAGAAAGT	ACCTATTTCA	ATGCTCTAGA	CAGCCTCAAA	TTTGAAAAAA
	851	ACACTTCTGG	GCAAGCTGGT	GGTGAATCT	ATACAGAGTC	TACGCTCACA
	901	ATCTCGAACA	TCACAAAATC	TATTGAATTT	ATCTCTAATA	AAGCTTCTGT
	951	CCCTGCCCCC	GCTCCTGAGC	CCACCTCTCC	GGCTCCAAGT	AGCTTAATAA
50	1001	ATTCTACAAC	GATCGATACC	TCGACTCTCC	AAACCCGAGC	AGCATCCGCA
	1051	ACTCCAGCAG	TGGCTCCTGT	TGCTGCCGTA	ACTCCAACAC	CAATCTCTAC
	1101	TCAAGAGACC	GCAGGAAATG	GAGGCGCTAT	CTATGCTAAA	CAAGGTATTT
	1151	CGATATCCAC	GTTTAAAGAT	CTGACCTTCA	AGTCTAAACT	TGCATCGGTA
	1201	GATGCCACCC	TTACTGTGCA	TTCTAGCACT	ATTGGAGAAT	CTGGAGGTGC
55	1251	TATCTTTGCA	GCAGACTCTA	TACAAATCCA	ACAGTGCACG	GGAACCACTT
	1301	TATTCAGTGG	CAATACTGCC	AATAAGTCTG	GTGGGGGTAT	TTACGCTGTA
	1351	GGACAAGTCA	CCCTAGAAGA	TATAGCGAAT	CTGAAGATGA	CCAACAACAC
	1401	CTGTAAAGGT	GAAGTGGGAG	CCATCTACAC	TAAAAGGCT	TTAACTATCA
	1451	ACAACGGTGC	CATTCTCACT	ACATTTTCTG	GAAATACATC	GACAGATAAT
60	1501	TCTGGGGCTA	TTTTTGCTGT	AGGTGGCATC	ACTCTCTCTG	ATCTTGTTAGA
	1551	AGTCCGCTTT	AGTAAAAATA	AGACCGGAAA	TTATTTCCGT	CCTATTACCA
	1601	AAGCGGCTAG	CAACACAGCT	CCTGTAGTTT	CTAGCTCTAC	AACTGCTGCA
	1651	TCTCCTGCGG	TCCCTGCTGC	CGCTGCAGCA	CCTGTTACAA	ACGCAGCAAA
	1701	AGTAGGGGCT	TTATATAGTA	CAGAAGGACT	GACTGTATCT	GGAATCACAT
65	1751	CGATATTGTC	GTTTGAAAAA	AACGAATGCC	AGAATCAAGG	AGGTGGGGCT

5

10

15

20

25

30

35

40

45

50

55

60

65

1801	TACGTTACTA	AAACCTTCCA	GTGTTCCGAT	TCTCATCGCC	TCCAGTTTAC
1851	TAGTAATAAA	GCAGCAGATG	AAGGCGGGGG	CCTGTATTGT	GGTGACGATG
1901	TCACGCTAAC	GAACCTGACA	GGGAAAACAC	TATTTCAAGA	GAATAGCAGT
1951	GAGAAACATG	GAGGTGGGCT	CTCTCTCGCC	TCAGGAAAAT	CTCTGACTAT
2001	GACATCGTTA	GAGAGCTTCT	GCTTAAATGC	AAATACAGCA	AAGGAAAACG
2051	GAGGCGGTGC	GAATGTCCCT	GAAAATATTG	TACTCACCTT	CACCTATACT
2101	CCCCTCCAA	ATGAACCTGC	GCCTGTGCAG	CAGCCCGTGT	ATGGAGAAGC
2151	TCTTGTACT	GGAAATACAG	CCACAAAAAG	TGGTGGGGGC	ATTTACACGA
2201	AAAAAGCGGC	CTTCTCAAAT	TTATCTTCTG	TAACTTTGA	TCAAAATACC
2251	TCCTCAGAAA	ATGGTGGTGC	CTTACTTACC	CAAAAAGCTG	CAGATAAAAC
2301	GGACTGTTCT	TTACCTATA	TTACAAATGT	CAATATCACC	AACAATACAG
2351	CTACAGGAAA	TGGTGGGGGC	ATTGCTGGGG	GAAAAGCACA	TTTCGATCGC
2401	ATTGATAATC	TTACAGTCCA	AAGCAACCAA	GCAAGAAAG	GTGGTGGGGT
2451	TTATCTTGAA	GATGCCCTCA	TCCTGGAAAA	GGTTATTACA	GGTTCTGTCT
2501	CACAAAATAC	AGCTACAGAA	AGTGGTGGGG	GTATCTACGC	TAAGGATATT
2551	CAACTACAAG	CTCTACCTGG	AAGCTTCACA	ATTACCGATA	ATAAAGTCGA
2601	AACTAGTCTT	ACTACTAGCA	CTAATTTATA	TGGTGGGGGC	ATCTATTCCA
2651	GTGGAGCTGT	CACGCTAACC	AATATATCTG	GAACCTTTGG	CATTACAGGA
2701	AACCTGTTA	TCAATACAGC	GACATCCAG	GATGCAGATA	TACAAGGTGG
2751	GGGCATTTAT	GCAACCACGT	CTCTCTCAAT	AAATCAATGT	AATACACCCA
2801	TTCTATTTAG	CAACAACCTCT	GCTGCCACTA	AAAAAACATC	AACAACAAAG
2851	CAAAATGCTG	GTGGGGCTAT	CTTCTCCGCT	GCAGTAACTA	TCGAGAATAA
2901	CTCTCAGCCC	ATTATTTTCT	TAAATAATTC	CGCAAAGTCG	GAAGCAACTA
2951	CAGCAGCAAC	TGCAGGAAAT	AAAGATAGCT	GTGGAGGAGC	CATTGCAGCT
3001	AACTCTGTTA	CTTTAACAAA	TAACCCTGAA	ATAACCTTTA	AAGGAAATTA
3051	TGCAGAAACT	GGAGGAGCGA	TTGGCTGTAT	TGATCTTACT	AATGGCTCAC
3101	CTCCCCGTAA	AGTCTCTATT	GCAGACAACG	GTTCGTCTCT	TTTTCAAGAC
3151	AACTCTGCGT	TAAATCGCGG	AGGCGCTATC	TATGGAGAGA	CTATCGATAT
3201	CTCCAGGACA	GGTGCAGCTT	TCATCGGTAA	CTCTTCAAAA	CATGATGGAA
3251	GTGCAATTTG	CTGTTCAACA	GCCCTAACTC	TTGCCCAAAA	CTCCCAACTT
3301	ATCTTTGAAA	ACAATAAGGT	TACGGAAACC	ACAGCCACTA	CAAAAGCTTC
3351	CATAAATAAT	TTAGGAGCTG	CAATTTATGG	AAATAATGAG	ACTAGTGACG
3401	TCACTATCTC	TTTATCAGCT	GAGAATGGAA	GTATTTTCTT	TAAAAACAAT
3451	CTATGCACAG	CAACAAACAA	ATACTGCAGT	ATTGCTGGAA	ACGTAAAATT
3501	TACAGCAATA	GAAGCTTCAG	CAGGGAAAGC	TATATCTTTC	TATGATGCAG
3551	TTAACGTTTC	CACCAAAGAA	ACAAATGCTC	AAGAGCTAAA	ATTAAATGAA
3601	AAAGCGACAA	GTACAGGAAC	GATTCCTATT	TCTGGGAAC	TTACGAAAAA
3651	TAAATCCTAT	ATTCCACAGA	AAGTCACTTT	CGCACATGGG	AATCTCATTC
3701	TAGGTAAAAA	TGCAGAACTT	AGCGTAGTTT	CCTTTACCCA	ATCTCCAGGC
3751	ACCACAATCA	CTATGGGCCC	AGGATCGGTT	CTTTCCAACC	ATAGCAAAGA
3801	AGCAGGAGGA	ATCGCTATAA	ACAATGTCAT	CATTGATTTT	AGTGAAATCG
3851	TTCTACTATA	AGATAATGCA	ACAGTAGCTC	CACCCACTCT	TAAATTAGTA
3901	TCGAGAACTA	ATGCAGATAG	TAAAGATAAG	ATTGATATTA	CAGGAAGTGT
3951	GACTCTTCTA	GATCCTAATG	GCAACTTATA	TCAAAATTCT	TATCTTGGTG
4001	AAGACCGCGA	TATCACTCTT	TTCAATATAG	ACAATCTGCG	AAGTGGGGCA
4051	GTTACAGCCA	CGAATGTCAC	CCTTCAAGGG	AAATTTAGGAG	CTAAAAAAGG
4101	ATATTTAGGA	ACCTGGAAAT	TGGATCCAAA	TTCTCGGGT	TCAAAAATTA
4151	TTCTAAAATG	GACCTTTGAC	AAATACCTGC	GCTGGCCCTA	CATCCCTAGA
4201	GACAACCACT	TCTACATCAA	CTCTATTTGG	GGAGCACAAA	ACTCTTTAGT
4251	GACTGTGAAA	CAAGGGATCT	TAGGGAACAT	GTTGAACAAT	GCAAGGTTTG
4301	AAGATCCTGC	TTTCAACAAC	TTCTGGGCTT	CGGCTATAGG	ATCTTTCTCT
4351	AGGAAAGAAG	TATCTCGAAA	TTCTGACTCA	TTACCTATAT	ATGGCAGAGG
4401	CTATACCGCT	GCTGTGGATG	CCAAACCTCG	CCAAGAATTT	ATTTTAGGAG
4451	CTGCCCTTCAG	TCAGGTTTTT	GGTCACGCCG	AGTCTGAATA	TCACCTTGAC
4501	AACTATAAGC	ATAAAGGCTC	AGGTCACTCT	ACACAAGCAT	CTCTTTATGC
4551	TGGCAATATC	TTCTATTTTC	CTGCGATACG	GTCTCGGCCT	ATTCATTATC
4601	AAGGTGTGGC	GACCTATGGT	TATATGCAAC	ATGACACCAC	AACCTACTAT
4651	CCTTCTATTG	AAGAAAAAAA	TATGGCAAAC	TGGGATAGCA	TTGCTTGGTT
4701	ATTTGATCTG	CGTTTCAGTG	TGGATCTTAA	AGAACCTCAA	CCTCACTCTA
4751	CAGCAAGGCT	TACCTTCTAT	ACAGAAGCTG	AGTATACCAG	AATTGCCCAG
4801	GAGAAATTC	CAGAGCTAGA	CTATGATCCT	AGATCTTTCT	CTGCATGCTC
4851	TTATGGAAAC	TTAGCAATTC	CTACTGGATT	CTCTGTAGAC	GGAGCATTAG
4901	CTTGGCGTGA	GATTATTTCTA	TATAATAAAG	TATCAGCTGC	GTACCTCCCT
4951	GTGATTTCTA	GGAATAATCC	AAAAGCGACC	TATGAAGTTC	TCTCTACAAA
5001	AGAAAAGGGC	AACGTAGTCA	ACGTTCTCCC	TACAAGAAAC	GCAGCTCGTG
5051	CAGAGGTGAG	CTCTCAAATT	TATCTTGGAA	GTTACTGGAC	ACTCTACGGC
5101	ACGTATACTA	TTGATGCTTC	AATGAATACT	TTAGTGCAAA	TGGCCAACGG
5151	AGGGATCCGG	TTTGTATTCT	AG		

The PSORT algorithm predicts an outer membrane location (0.926).

The protein was expressed in *E.coli* and purified as a GST-fusion (Figure 60A) or his-tagged product. The recombinant proteins were used to immunise mice, whose sera were used in Western blot (Figure 60B) and FACS (Figure 60C) analyses.

- 5 The cp6830 protein was also identified in the 2D-PAGE experiment (Cpn0540) and showed good cross-reactivity with human sera, including sera from patients with pneumonitis.

These experiments show that cp6830 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

#### Example 61

- 10 The following *C.pneumoniae* protein (PID 4376854) was expressed <SEQ ID 121; cp6854>:

```

1  MSIAIAREQY AAILDMHPKP SIAMFSSEQA RTSWEKROAH PYLYRLLEII
51  WGVVFKFLGL IFFIPLGLFW VLQKICQNF I LLGAGGWIFR PICRDSNLLR
101 QAYAARLFSA SFQDHVSSVR RVCLQYDEVF IDGLELRLPN AKPDRWMLIS
151 NGNSDCLEYR TVLQGEKDWI FRIAESQSN ILIFNYPGVM KSGGNITRNN
15  201 VVKSQYACVR YLRDEPAGPQ ARQIVAYGYS LGASVQAEAL SKEIADGSDS
251 VRWFVVKDRG ARSTGAVAKQ FIGSLGVWLA NLTHWNINSE KRSKDLHCPE
301 LFIYKDSQG NLIGDGLFKK ETCFAAPFLD PKNLEBCSGK KIPVAQTGLR
351 HDHILSDDVI KEVAGHIQRH PDN*
```

The cp6854 nucleotide sequence <SEQ ID 122> is:

```

20 1  ATGTCAATAG CTATTGCAAG GGAACAATAC GCAGCTATAT TGGATATGCA
51  TCCTAAACCT TCGATCGCCA TGTTTTCTTC GGAGCAGGCG AGAACTTCTT
101 GGGAGAAACG ACAGGCTCAT CCTTACCTTT ATCGTCTTCT TGAGATCATA
151 TGGGGTGTG TGAAATTTCT TCTCGGCTTA ATCTTCTTTA TTCCCTTGGG
20 201 TCTTTTCTGG GTCCTTCAGA AGATAATGTC GAATTTTATT CTCTTTGGTG
25 251 CAGGAGGGTG GATTTTTFAGA CCCATATGCA GGGACTCTAA TTTATTGCGA
301 CAAGCTTACG CCGCGCGTCT TTTCTCCGCT TCATTCCAAG ATCATGTCTC
351 CTCTGTGCGA AGGGTTTGCT TACAGTATGA CGAGGTCTTT ATTGACGGAT
401 TGGAGTTACG TCTTCCCAAT GCTAAGCCAG ATCGATGGAT GTTAATCTCC
451 AATGGAAACT CCGATTGCTT AGAGTATAGG ACAGTGCTGC AAGGGGAAAA
30 501 GGAAGTGGTA TTCCGTATTG CTGAAGAGTC TCAATCCAAC ATTTTAAATCT
551 TCAATTACCC AGGAGTCATG AAGAGCCAAG GGAATATAAC AAGAAACAAT
601 GTAGTCAAAT CTTATCAAGC ATGCGTACGC TATCTTAGAG ATGAACCCGC
651 AGGACCTCAG CCGCGTCAAA TCGTTGCTTA TGGCTATTCT TTAGGAGCTA
701 GTGTTCAAGC CGAAGCATT AAGTAAAGAGA TCGCAGACGG AAGTGATAGC
35 751 GTCCGTGGT TTGTCGTAA AGATCAGAGG GCTCGCTCTA CAGGAGCCGT
801 TGCTAAACAG TTTATTGGAA GTCTAGGAGT TTGGCTGGCG AATCTTACCC
851 ATTGGAATAT TAATCTTGAA AAGAGAAGCA AGGACTTGCA TTGCCAGAA
901 CTCTTTATTT ATGGCAAGGA TTCCCAAGGT AATCTTATCG GGGATGGATT
951 GTTCAAAAAA GAGACGTGCT TCGCAGCACC ATTTTATAGT CCTAAAAACT
40 1001 TGGAAGAGTG TTCAGGGAAG AAAATCCCTG TAGCTCAGAC CGGTCTAAGA
1051 CACGATCATA TCCTTTCCGA TGATGTGATT AAAGAAGTTG CAGGTCATAT
1101 TCAAGACAT TTCGATAATT A
```

The PSORT algorithm predicts an inner membrane location (0.461).

The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 61A.

- 45 The recombinant protein was used to immunise mice, whose sera were used in Western blot (Figure 61B) and FACS (Figure 61C) analyses. A his-tagged protein was also expressed.

These experiments show that cp6854 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

## Example 62

The following *C.pneumoniae* protein (PID 4377101) was expressed <SEQ ID 123; cp7101>:

```

1  MYSCYSKGIS HNYLLHPMSR LDIFVFDLSI ANQDQNLLEE IFCSEDTVLF
51  KAYRTTALQS PLAANKNLNIA RKVANYILAD NGEIDTVKLV EAIHHLSQCT
101 YPLGPHRHNE AQDREHLLKM LKALKENPKL KESIKTLFVP SYSTIQNLIR
151 HTLALNPQTI LSTIHVRQAA LTALFTYLRQ DVGSCFATAP AILIHQBYPE
201 RFLKDLNDLI SSGKLSRIVN QREIAPVINL SGCIGELFKP LRILDLYPDP
251 LVKLSSSPGL KKAFFSAANLI ETLGDSEAQI QQLLSHQYLM QKLQNVHETL
301 TANDIIKSTL LHYYQLQEST VRAIFFKEGL FSKEQVAFST QHPRELSBIQ
351 RYVHYLHAYE EAKSAFIHDT QNPLLKAWAY TLATLADASQ PTISNHIRLA
401 LGWKSSEDPHS LVS LVTHFVE EEVENIRILV QQCEQTYHEA RSQLEYIEGR
451 MRNPLNNQDS QILTMHMRF RQELNKALYE WDSAQEKAKK FLHLPEFLLS
501 FYTQKIPLYF RSSYDAFIQE FAHLYANAPA GFRILFTHGR THPNTWSPYI
551 SINEFIRFLS EFFTSTESL LGKHAVINLE KETSRLVHNI TAMLHTDVFO
601 EALLTRILEA YQLPVPPSIL NHL DQLSQTP WVYVSGGTV D TLLLDYFESS
651 EPLTLTEKHP ENPHELAAPY ADALKDLPTG IKSYLEEGSH SLLSSSPHTV
701 FSIIAGSPLF REAWDNDWYS YTWLRDVWVK QHQDFLODTI LPQLSIYAFI
751 ENFCNKYALQ HVVHDFHDFC SDHSLTLPEL YDKGSRFLSS LFTKDKTVAL
801 IYIRLLLYLM VREVFPVSEQ QLPEVLDNVS SYLGISSRIT YEKFRSLIEE
851 TIPKMTLLSS ADLRHIYKGL LMQSYQKIYT BEDTYLRLLT AMRHHNLAYP
901 APLLFADSNW PSYFPGFILN PGTTEIDLWK FNYAGLQGGP LDNIQELFAT
951 SRPWTLYANP IDYGMPPPPG YRSRLPKEFF *
```

The cp7101 nucleotide sequence <SEQ ID 124> is:

```

1  ATGTATTCGT GTTACAGCAA AGGAATATCC CATAACTATC TTCTACATCC
25 51  TATGTCACGT TTGGATATTT TTGTTTTCGA TTCTCTGATC GCAAACCAGG
101 ATCAAATCT TCTTGAGGAA ATTTTCTGTT CTGAAGACAC AGTTTATTAT
151 AAAGCCTACC GTACTACGGC TCTACAATCC CCTCTAGCTG CTAAGAACCT
201 AAATATCGCC CGTAAAGTCG CAAATTATAT CTTAGCTGAC AATGGGGAAA
251 TCGATACAGT AAAGCTTGTC GAAGCCATTC ACCATCTCTC ACAATGTACC
301 TATCCTTTAG GGCCTCATCG CCATAATGAA GCTCAAGATC GTGAACACCT
351 CCTTAAATG CTAAGAGCTC TAAAGGAAA TCCTAAATTA AAAGAAAGCA
401 TCATAACTCT CTTTGTCCCT TCATACTCTA CAATCCAAAA CCTAATTCGC
451 CATACTACTAG CATTTGAATCC ACAGACAATT CTCTCTACGA TTCATGTGCG
501 TCAAGCAGCA CTCACAGCGC TCTTCACCTA CCTTCGGCAA GATGTAGGTT
35 551 CCTGTTTTCG TACGGCTCCT GCCATTCTCA TTCACCAAGA ATATCCAGAA
601 CGATTCTCTA AAGATCTCAA TGATCTCATT AGCAGTGGCA AACTCTCTAG
651 AATCGTAAAC CAAAGGGAAA TTGCGGTTCC TATAAACCTT TCGGGATGCA
701 TTGGAGAGCT ATTCAAGCCT TTAAGGATTC TAGATCTTTA TCCTGATCCT
751 CTGGTTAAGC TCTCTCATC TCCAGGACTC AAAAAAGCCT TTTCTGCTGC
40 801 CAATCTTATT GAAACTCTTG GGGATTCTGA AGCACAAATC CAACAGTTGC
851 TCTCGCATCA ATATTGATG CAAAAACTAC AAAATGTCCA TGAGACCTTA
901 ACTGCTAAGC ACATTATCAA ATCGACACTT CTGCACTACT ATCAGCTCCA
951 AGAAAGTACT GTACGAGCTA TTTCTTCAA AGAAGGGTTG TTCAGCAAAG
1001 AACAAGTGGC ATTCTCGACG CAACACCCCA GAGAGCTCTC AGAAATACAA
45 1051 CGGGTATACC ACTACTTACA TGCCTATGAA GAAGCAAAAT CTGCTTTTAT
1101 CCATGACACT CAAAATCCCT TACTGAAAGC CTGGGAGTAT ACTTTAGCGA
1151 CTCTTGCGGA TGCTAGCCAA CCTACCATCT CAAACCATAT CCGCCTTGCC
1201 TTAGGATGGA AAAGTGAAGA CCCTCACAGT CTTGTATCTC TAGTTACACA
1251 CTTTGTGAA GAGGAAGTAG AAAACATCCG AATTTTAGTC CAACAATGTG
50 1301 AACAGACCTA TCACGAAGCA CGCTCCCAAC TAGAATATAT TGAAGGGCGG
1351 ATGCGCAACC CACTAAATAA TCAAGACAGT CAGATTTTGA CGATGGATCA
1401 CATGCGCTTC CGTCAAGAAC TCAATAAAGC TCTTTATGAG TGGGATAGTG
1451 CTCAAGAAAA GGCAAGAAA TTTCTACATC TTCCTGAATT CTTACTTTCT
1501 TTTCTATACAA AGCAAAATTC CTTATACTTT CGTAGTTCTT ACGATGCCTT
55 1551 CATTCAAGAA TTTGCTCATC TCTATGCTAA TGCTCCCGCT GGCTTCCGTA
1601 TTTCTTTTAC GCATGGACGC ACCCATCCGA ACACATGGTC CCCCATCTAT
1651 TCGATTAAATG AATTTATACG TTTTCTTTCT GAATTTCTCA CCTCCACAGA
1701 GTCAGAACTT TCGGGGAAAC ATGCCGTGAT CAATTTAGAG AAAGAAACAT
1751 CTCGGCTCGT CCACAACATC ACTGCCATGC TACACACGGA TGTTTTCCAA
60 1801 GAAGCTCTCC TTACAAGAA TTTAGAAGCC TATCAGCTTC CTGTGCCTCC
1851 CTCCATCTTA AACCACCTAG ATCAGCTGTC ACAAACCTCC TGGGTTTATG
1901 TTTCTGGAGG AACAGTGGAC ACTCTTCTTT TGGATTATTT TGAAAGCTCA
1951 GAACCTCTGA CACTTACAGA AAAGCATCCT GAAAATCCTC ATGAGCTTGC
2001 AGCTTTCTAC GCAGACGCCC TTAAAGATCT CCCTACAGGA ATTAAAGT
```

-103-

```

2051 ATCTAGAAGA AGGATCCCAC TCTCTACTTA GCTCATCACC CACCCACGTT
2101 TTCTCTATAA TCGCAGGATC TCCTTTATTT CGGGAAGCTT GGGATAATGA
2151 TTGGTACAGC TATACCTGGC TTCGTGATGT CTGGGTGAAA CAACACCAAG
2201 ATTTCCCTTCA AGATACTATA TTACCTCAGC TAAGTATCTA TGCTTTCATA
5 2251 GAGAATTTT GTAAACAAATA TGCTTTGCAA CATGTAGTTC ATGACTTTCA
2301 TGATTCTGCT TCCGACCACT CCTTGACTCT TCCGGAGCTC TATGACAAAG
2351 GATCGCGTTT TCTAAGCTCC TTATTCACCA AAGATAAGAC CGTAGCTCTT
2401 ATCTATATAC GCCGTCTTCT CTACCTTATG GTCCGTGAAG TCCCTTATGT
2451 TTCAGAACAA CAGCTTCCAG AAGTCTTAGA TAACGTCTCT TCATATCTCG
10 2501 GGATTTCCTC TCGTATTACC TATGAGAAAT TCCGCTCCCT GATAGAGGAA
2551 ACCATCCCTA AAATGACCTT ACTCTCCTCA GCAGACCTGA GGCATATCTA
2601 TAAAGGTCTC CTCATGCAAA GTTATCAAAA GATCTACACC GAAGAAGATA
2651 CGTACCTCCG CCTCACCACG GCAATGAGGC ATCATAATCT TGCCATATCCC
2701 GCTCCTTTGC TCTTTGCAGA CAGTAACTGG CCTTCTATTT ATTTTGGATT
15 2751 CATCTTAAAT CCAGGAACCA CAGAGATCGA TCTTTGGAAA TTTAACTATG
2801 CAGGGCTGCA AGGACAGCCT CTTGACAATA TCCAGGAGCT GTTCGCAACG
2851 TCAAGACCCT GGACCCTCTA TGCAAATCCT ATAGATTATG GCATGCCACC
2901 GCCTCCAGGC TACCGCAGCC GCCTCCCTAA AGAATTTTTC TAG

```

The PSORT algorithm predicts a cytoplasmic location (0.206).

- 20 The protein was expressed in *E.coli* and purified as a GST-fusion (Figure 62A) or his-tagged product. The proteins were used to immunise mice, whose sera were used in Western blot (Figure 62B) and FACS (Figure 62C) analyses.

This protein also showed good cross-reactivity with human sera, including sera from patients with pneumonitis.

- 25 These experiments show that cp7101 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 63

The following *C.pneumoniae* protein (PID 4377107) was expressed <SEQ ID 125; cp7107>:

```

30 1 MSIVRNSALP LPCLSRSETF KKVRSHEMKFM KVLTPWIYRK DLWVTAFLLT
51 AIPGSFAHTL VDIAGEPRHA AQATGVSGDG KIVIGMKVPD DFFAITVGFQ
101 YIDGHLQPLE AVRPQCSVYP NGITPDGTVI VGTNYAIGMG SVAVKWVNGK
151 VSELPMLPDT LDSVASAVSA DGRVIGGNRN INLGASVAVK WEDDVITQLP
201 SLPDAMNACV NGISSDGSII VGTMDVSWR NTAVQWIGDQ LSVIGTLGGT
251 TSVASAISTD GTVIVGGSSE ADSQTHAYAY KNGVMSDIGT LGGFYSLAHA
35 301 VSSDGSVIVG VSTNSEHRYH AFQYADGQMV DLGTLGGPES YAQGVSGDGK
351 VIVGRAQVPS GDWHAFPCPF QAPSPAPVHG GSTVVTSQNP RGMVDINATY
401 SSLKNSQQQL QRLLIQHSAK VESVSSGAPS FTSVRGALSK QSPAVQNDVQ
451 KGTFLSYRSQ VHGNVQNOQL LTGAFMDWKL ASAPKCGFKV ALHYGSQDAL
501 VERAALPYTE QGLGSSVLSG FGGQVQGRYD FNLGETVVLQ PFMGIQVLHL
40 551 SREGYSEKNV RFPVSYDSVA YSAATSFMGA HVFASLSPKM STAATLGVER
601 DLNSHIDEFK GSVSAMGNFV LENSTVSVLR PFASLAMYVD VRQQQLVLTLS
651 VVMNQQLTGT TSLVLSQSSY NLSF*

```

The cp7107 nucleotide sequence <SEQ ID 126> is:

```

45 1 ATGAGTATAG TCAGAAATTC TGCATTGCCA CTTCCGTGTT TAAGCAGATC
51 CGAAACCTTT AAAAAAGTTA GGTGCGCATAT GAAATTTATG AAAGTCCTTA
101 CTCCATGGAT TTATCGAAAA GATCTTTGGG TAACAGCATT CTTACTGACA
151 GCAATTCAG GATCTTTTGC ACATACTCTT GTTGATATAG CAGGAGAACC
201 TCGGCATGCT GCTCAAGCAA CAGGAGTTTC TGGAGATGGT AAAATTGTTA
251 TAGGAATGAA AGTTCCGGAT GATCCTTTTG CTATAACTGT AGGATTTCAA
50 301 TATATTGATG GGCATTTGCA ACCCTTAGAG GCAGTACGTC CTCAATGCTC
351 TGTATACCCT AATGGTATAA CCCCAGACGG AACGGTTATT GTGGGTACAA
401 ACTATGCCAT CGGGATGGGT AGTGTGTGCTG TGAAATGGGT AAATGGCAAG
451 GTTCTTGAAC TTCCCATGCT CCTGACACC CTCGATTCTG TAGCATCGGC
501 AGTTTCTGCA GATGGAAGAG TGATTGGAGG GAATAGAAAT ATAAATCTTG
55 551 GCGCTTCTGT TGCTGTGAAA TGGGAGGACG ACGTGATTAC ACAACTTCCT
601 TCTCTTCTCG ATGCTATGAA TGCTTGTGTT AACGGAATTT CTTCAGATGG

```

-104-

5 651 TTCTATAATT GTAGGAACCA TGGTAGACGT GTCATGGAGA AATACCGCAG  
 701 TACAATGGAT CGGGGATCAG CTCTCTGTTA TTGGGACTTT AGGAGGAACT  
 751 ACTTCTGTFTG CTAGTGCAAT CTCAACAGAT GGCACGTGTGA TTGTAGGAGG  
 801 TTCTGAAAAT GCAGATTCTC AGACTCATGC CTATGCTTAT AAAAACGGTG  
 851 TTATGAGCGA TATAGGGACC CTCGAGGTT TTTATCTTT AGCACATGCA  
 901 GTATCTTCAG ATGGTCTGTG GATTGTAGGA GTATCCACGA ACTCTGAGCA  
 951 TAGATATCAT GCATFCCAAT ATGCTGATGG ACAGATGGTA GATTTAGGAA  
 1001 CTTTAGGAGG GCCTGAATCT TATGCTCAAG GTGTGTCTGG AGATGGAAAG  
 1051 GTAATTGTGG GTAGAGCACA AGTACCATCT GGAGATTGGC ATGCGTTCCT  
 1101 ATGTCCTTTC CAAGCTCCGA GCCCTGCTCC TGTCCATGGG GGAAGCACTG  
 1151 TCGTAACTAG CCAGAATCCA CGTGAATGG TAGATATCAA TGCTACGTAC  
 1201 TCCTCTTTGA AAAATAGCCA ACAACAATA CAAAGATTGC TTATCCAGCA  
 1251 TAGTGCAAAA GTTGAAAGTG TATCCTCAGG AGCACCATCT TTTACAAGTG  
 1301 TGAAGAGTGC GATCTCAAAA CAGAGCCCTG CAGTGCAAAA TGATGTACAG  
 1351 AAAGGGACGT TTTTAAAGTTA CCGTCCCAA GTTCATGGAA ACGTGCAGAA  
 1401 TCAGCAATTG CTCACAGGAG CTTTATGGA CTGGAAACTC GCTTCAGCTC  
 1451 CTAAATGCGG CTTTAAAGTA GCTCTCCACT ATGGCTCTCA AGATGCTCTC  
 1501 GTAGAACGTG CAGCTCTTCC TTACACAGAA CAAGGCTTAG GAAGCAGTGT  
 1551 CTTGTTCAGGT TTTGGAGGAC AAGTTCAGG ACCTATGAC TTTAATTAG  
 20 1601 GAGAACTGT TGTCTGCAA CCTTTATGG GCATTCAAGT TCTCCACCTA  
 1651 AGTAGAGAAG GGTATCTGA GAAGAATGTT CGATTTCCTG TAAGCTATGA  
 1701 TTCTGTAGCC TACTCAGCAG CTTACTAGCTT TATGGGTGCG CATGTATTTG  
 1751 CCTCCCTAAG CCTTAAATG AGTACAGCAG CAACTTTAGG TGTGGAGAGA  
 1801 GATCTGAATT CACATATAGA TGAATTTAAG GGATCCGTCT CTGCTATGGG  
 25 1851 AAACTTTGTG TTGAAAAT CTACAGTGAG TGTTTTAAAGA CCTTTTGCTT  
 1901 CTCTTGCTAT GTACTATGAC GTAAGACAAC AGCAACTCGT GACGTTGTCA  
 1951 GTAGTTATGA ATCAACAACC CTTAACAGGC ACCTAAGCT TAGTAAGCCA  
 2001 AAGTAGCTAT AATCTTAGCT TCTAA

The PSORT algorithm predicts an inner membrane location (0.100).

30 The protein was expressed in *E.coli* and purified as a GST-fusion (Figure 63A) or his-tagged product. The proteins were used to immunise mice, whose sera were used in Western blot (Figure 63B) and FACS (Figure 63C) analyses.

These experiments show that cp7107 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### 35 Example 64

The following *C.pneumoniae* protein (PID 4376467) was expressed <SEQ ID 127; cp6467>:

40 1 MLRFFAVFIS TLWLITSGCS PSQSSKGIFV VNMKEMPRSL DPGKTRLIAD  
 51 QTLMRHLYEG LVEEHSQNGE IKPALAESYT ISEDTRYTF KIKNILWSNG  
 101 DPLTAQDFVS SWKEILKEDA SSVLYAFLP IKNARAIFDD TESPENLGVR  
 151 ALDKRHLEIQ LETPCAHLFLH FLTLPIFFPV HETLRNYSYS FEEMPITCGA  
 201 FRPVSLEKGL RLHLEKNPMY HNKSrvKLHK IIVQFISNAN TAAILFKHKK  
 251 LDWQGPWPGE PIPPEISASL HQDDQLFSLP GASTWLLFN IQKKPWNNAK  
 301 LRKALSLAID KDMLTKVVYQ GLAEPTDHIL HPRLYPGTYP ERKRQNERIL  
 351 EAQQLFEEAL DELQMTREDL EKETLTFSTF SFSYGRICQM LRQWKKVLK  
 45 401 FTIPIVGQEF FTIQKNFLEG NYSLVFNQWT AAFIDPMSYL MIFANPGGIS  
 451 PYHLQDSHFQ TLLIKITQEH KKHLRNQLII BALDYLEHCH ILEPLCHPNL  
 501 RIALNKNIKN FNLFVRRTSD FRFIEKL\*

A predicted signal peptide is highlighted.

The cp6467 nucleotide sequence <SEQ ID 128> is:

50 1 ATGCTCCGTT TCTTCGCTGT ATTTATATCA ACTCTTTGGC TCATTACCTC  
 51 AGGATGTTCC CCATCCCAAT CCTCTAAAGG AATTTTGTG GTAAATATGA  
 101 AGGAAATGCC ACGCTCCTTG GATCCTGGAA AAACCTCGTCT CATTGCAGAC  
 151 CAACTCTAA TGGCTCATCT ATATGAAGGA CTCGTCGAAG AACATTCCCA  
 201 AAATGGAGAG ATTAAACCAG CCCTTGCAGA AAGCTACACC ATCTCCGAAG  
 55 251 ACGGGACTCG GTACACATT AAAATCAAAA ACATCCTTTG GAGTAACGGA  
 301 GACCCTCTGA CAGCTCAAGA CTTTGTCTCC TCTTGAAGG AAATCCTAAA

351 GGAAGATGCG TCCTCCGTAT ATCTCTATGC GTTTTACCT ATCAAAAATG  
 401 CTCGGGCAAT CTTTGATGAT ACTGAGTCTC CAGAAAATCT AGGAGTCCGA  
 451 GCTTTAGATA AGCGTCATCT CGAAATTCAG TTAGAAACTC CCTGCGCGCA  
 501 TTTCCTACAT TTCTTGACTC TTCTTATTTT TTCCCTGTT CATGAAACTC  
 551 TGCGAAACTA TAGCACCTCT TTTGAAGAGA TGCCCATAC CTGCGGTGCT  
 601 TTCCGCCCTG TGTCTCTAGA AAAAGGCCTG AGACTCCATC TAGAGAAAAA  
 651 CCCTATGTAC CATAATAAAA GCCGTGTGAA ACTACATAAA ATTATTGTAC  
 701 AGTTTATCTC AAACGCTAAC ACTGCAGCCA TTCTATTCAA ACATAAGAAA  
 751 TTAGATTGGC AAGGACCTCC TTGGGGAGAA CCTATCCCTC CAGAAATCTC  
 801 AGCTTCTCTA CATCAAGATG ACCAGCTCTT TTCTCTTCCG GCGCTTCGA  
 851 CTACATGGTT ACTCTTTAAT ATACAAAAAA AACCTTGGA CAATGCTAAA  
 901 TTACGCAAGG CATTGAGCCT TGCAATAGAC AAAGATATGT TAACCAAAGT  
 951 GGTATACCAA GGTCTTGCAG AACCTACAGA TCATATCCTA CATCCAAGAC  
 1001 TTTATCCAGG GACCTATCCC GAACGAAAAA GACAAAACGA AAGAAATCTT  
 1051 GAGGCTCAAC AACTCTTTGA AGAAGCTCTA GACGAACTTC AAATGACACG  
 1101 CGAAGATCTA GAAAAGGAAA CTTTGACTTT CTCAACCTTT TCTTTTCTT  
 1151 ACGGAAGGAT TTGCCAAATG CTAAGAGAAC AATGGAAGAA AGTCTTAAAA  
 1201 TTTACTATCC CTATAGTAGG CCAAGAGTTT TTCACAATAC AAAAAAATTT  
 1251 CCTAGAGGGG AACTATTCCC TAACCGTGAA CCAATGGACC GCAGCATTTA  
 1301 TTGATCCGAT GTCTTATCTC ATGATCTTTG CCAATCCTGG AGGAATTTCC  
 1351 CCCTATCACC TCCAAGATTC ACACTTTCAA ACTCTTCTCA TAAAGATCAC  
 1401 TCAAGAACAT AAAAAACACC TACGAAATCA GCTTATATT GAAGCCCTTG  
 1451 ACTATTTAGA AACTGTCTAC ATTCTCGAAC CACTATGTCA TCCAAATCTT  
 1501 CGAATTGCTT TGAACAAAAA CATTAAAAAC TTTAATCTTT TTGTTGACG  
 1551 AACTTCAGAC TTTCTTTTAA TAGAAAAACT ATAG

The PSORT algorithm predicts an outer membrane lipoprotein (0.790).

The protein was expressed in *E.coli* and purified as a his-tag product and a GST-fusion protein, as shown in Figure 64A. The recombinant his-tag protein was used to immunise mice, whose sera were used in a Western blot (Figure 64B). The recombinant GST-fusion protein was also used to immunise mice, whose sera were used in a Western blot (Figure 64C) and for FACS analysis (Figure 64D).

These experiments show that cp6467 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

#### Example 65

The following *C.pneumoniae* protein (PID 4376679) was expressed <SEQ ID 129; cp6679>:

1 MRKMLVLLAS LGLLSPTLSS CTHLGSSGSY HPKLYTSGSK TKGVIAMLPV  
 51 FHRPGKSLEP LPWNLQGEFT KEISKRFYAS EKVFLIKHNA SPQTVSQFYA  
 101 PIANRLPETI IEQFLPAEFI VATELLEQKT GKEAGVDSVT ASVRVRVFDI  
 151 RHHKIALIYQ EIIECSQPLT TLVNDYHRYG WNSKHFDSTP MGLMHSRLFR  
 201 EVVARVEGYV CANYS\*

A predicted signal peptide is highlighted.

The cp6679 nucleotide sequence <SEQ ID 130> is:

1 ATGCGAAAAA TGTGCTATT ATTGGCATCT TTAGGACTTC TATCCCCAAC  
 51 CCTATCCAGC TGCACCTACT TAGGCTCTTC AGGAAGTTAT CATCCTAAGC  
 101 TATACACTTC AGGGAGCAAA ACTAAAGGTG TGATTGCGAT GCTTCTGTGA  
 151 TTTCATCGCC CAGGAAAGAG TCTTGAACCT TTACCTTGGA ACCTCCAAGG  
 201 AGAATTTACT GAAGAGATCA GCAAAAGGTT TTATGCTTCG GAAAAGGTCT  
 251 TCCTGATCAA GCACAATGCT TCACCTCAGA CAGTCTCTCA GTTCTATGCT  
 301 CCGATTGCGA ATCGTCTACC CGAAACAATT ATTGAGCAAT TTCTTCCTGC  
 351 AGAATTCATT GTTGCTACAG AACTGTTAGA ACAAAGACA GGGAAAGAAG  
 401 CAGGTGTCGA TTCTGTAACA GCGTCTGTAC GTGTTCGCGT TTTTGATATC  
 451 CGTCATCATA AAATAGCTCT CATTATCAA GAGATTATCG AATGCAGCCA  
 501 GCCTTTAACT ACCCTAGTCA ATGATTATCA TCGCTATGGC TGGAACTCAA  
 551 AACATTTTGA TTCAACGCCC ATGGGCTTAA TGCATAGCCG TCTTTTCCGC



601 GAAGTTGTTG CCAGAGTTGA GGGCTATGTT TGTGCTAACT ACTCGTAG

The PSORT algorithm predicts an inner membrane location (0.149).

The protein was expressed in *E.coli* and purified as a his-tag product (Figure 65A) and as a GST-fusion product (Figure 65B). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 65C) and for FACS analysis.

These experiments show that cp6679 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

#### Example 66

The following *C.pneumoniae* protein (PID 4376890) was expressed <SEQ ID 131; cp6890>:

```

10      1  MKQLLFCVCV FAMSCSAYAS PRRQDPSVMK ETRNNYGI VSGQEWVVRG
      51  SDGTTTKVLK NGATLHEVYS GLLHGEITL TFPHTALDV VQIYDQGRV
      101 SRKTFPVNGL PSQELFNEF GTFVLTRWPD NNSDTITKP YFIETTYQGH
      151 VIEGYSYTSFN GKYSSSIHNG EGVRSVFSSN NILLSEETFN EGVMVKYTTF
      201 YPNRDPESIT HYQNGQPHGL RLTYLQGGIP NTIEEWRYGF QDGTITVFKN
      251 GCKTSEIAYV KGVKEGLELR YNBQEIVAEB VSWRNDFLHG ERKIYAGGIQ
      301 KHEWYRGRS VSKAKFERLN AAG*

```

A predicted signal peptide is highlighted.

The cp6890 nucleotide sequence <SEQ ID 132> is:

```

20      1  ATGAAACAAT TACTTTTCTG TGTTCGCGTA TTTGCTATGT CATGTTCTGC
      51  TTACGCATCC CCACGACGAC AAGATCCTTC TGTATGAAG GAAACATTCC
      101 GAAATAATTA TGGCATTATT GTTCCGGTC AAGAATGGGT AAAGCGTGGT
      151 TCTGACGGCA CCATCACCAA AGTACTCAA AATGGAGCTA CCCTGCATGA
      201 AGTTTATTCT GGAGGCCTCC TTCATGGGGA AATTACCTTA ACGTTTCCCC
      251 ATACCACAGC ATTGGACGTT GTTCAAATCT ATGATCAAGG TAGACTCGTT
      301 TCTCGCAAAA CCTTTTGTGT GAACGGTCTT CCATCTCAAG AAGAGCTGTT
      351 CAATGAAGAT GGCACGTTTG TCCTCACACG ATGGCCGGAC AACACGACA
      401 GTGATACCAT CACAAAGCCT TACTTCATAG AAACGACATA TCAAGGCAT
      451 GTCATAGAAG GAAGTTATAC TTCCTTTAAT GGGAAATACT CCTCATCCAT
      501 CCACATGGA GAGGGAGTTC GTTCTGTGTT CTCTCCAAT AACATCCTTC
      551 TTTCTGAAGA GACCTTCAAT GAAGGTGTCA TGGTGAAATA TACCACATTC
      601 TATCCGAATC GCGATCCCGA ATCGATTACT CATTATCAAA ATGGACAGCC
      651 TCACGGCTTA CGGCTAACAT ATCTACAAGG TGGCATCCCC AATACGATAG
      701 AGGAGTGGCG TTATGGCTTT CAAGACGGAA CGACCATCGT ATTTAAAAAT
      751 GGTGTGAAGA CATCTGAGAT CGCTTATGTT AAGGGAGTGA AAGAAGGTTT
      801 AGAACTGCGC TACAATGAAC AGGAAATTGT AGCTGAAGAA GTTCTCTGGC
      851 GTAATGATTT TCTGCATGGA GAACGTAAGA TCTATGCTGG AGGAATCCAA
      901 AAGCATGAAT GGTATTACCG CGGGAGATCT GTATCTAAAG CCAAATTCGA
      951 GCGGCTAAAT GCTGCAGGAT AG

```

The PSORT algorithm predicts an outer membrane location (0.940).

The protein was expressed in *E.coli* and purified as a GST-fusion product, as shown in Figure 66A. The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 66B) and for FACS analysis. A his-tagged protein was also expressed.

These experiments show that cp6890 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

#### Example 67

The following *C.pneumoniae* protein (PID 6172323) was expressed <SEQ ID 133; cp0018>:

-107-

1 MKTSVSMLLA LLCSGASSIV LHAATTPPLNP EDGFIGEGNT NTFSPKSTTD  
 51 AAGTTYSLTG EVLYIDPGKG GSI TGTCTFVE TAGDLTFLGN GNTLKFLSVD  
 101 AGANLAVAHV QGSKNLSFTD FLSLVITESP KSAVTTGKGS LVSLGAVQLQ  
 151 DINTLVLTSTN ASVEDGGVIK GNSCLIQGIK NSAI FGQNTS SKKGGAISTT  
 5 QGLTIENNLG TLKFNENKAV TSGGALDLGA ASTFTANHEL IFSQNKTSGN  
 201 AANGGAINCS GDLTFTDNTS LLLQENSTMQ DGGALCSTGT ISITGSDSIN  
 251 VIGNTSGQKG GAISAASLKI LGGQGGALFS NNVVTHATPL GGAIFINTGG  
 301 SLQLFTQGGD IVFEGNQVTT TAPNATTKRN VIHLESTAKW TGLAASQGNA  
 351 IYFYDPITTN DTGASDNLRI NEVSANQKLS GSIVFSGERL STAEALAEENL  
 10 401 TSRLNQPVTL VEGSLVLRQG VTLLITQGFSSQ EPESTLLLDL GTSL\*  
 451

A predicted signal peptide is highlighted.

The cp0018 nucleotide sequence <SEQ ID 134> is:

1 ATGAAGACTT CAGTTTCTAT GTTGTGGGCC CTGCTTTGCT CGGGGGCTAG  
 51 CTCTATTGTA CTCCATGCCG CAACCACTCC ACTAAATCCT GAAGATGGGT  
 15 101 TTATTGGGGA GGCAATACA AATACCTTTT CTCCGAAATC TACAACGGAT  
 151 GATTGCAGGA CTACCTACTC TCTCACAGGA GAGGTCTGT ATATAGATCC  
 201 GGGGAAAGGT GGTCAATTA CAGGAACTTG CTTTGTAGAA ACTGCTGGCG  
 251 ATCTTACATT TTTAGGTAAT GGAAATACCC TAAAGTTCCT GTCGGTAGAT  
 301 GCAGGTGCTA ATATCGCGGT TGCTCATGTA CAAGGAAGTA AGAATTTAAG  
 20 351 CTTTCACAGAT TTCTTTCTC TGGTGATCAC AGAATCTCCA AAATCCGCTG  
 401 TTACTACAGG AAAAGGTAGC CTAGTCAGTT TAGGTGCAGT CCAACTGCAA  
 451 GATATAAACA CTCTAGTTCT TACAAGCAAT GCCTCTGTCG AAGATGGTGG  
 501 CGTGATTAAA GAAACTCCT GCTTGATTCA GGGAAATCAA AATAGTGCAG  
 551 TTTTGGACA AAATACATCT TCGAAAAAG GAGGGCGAT CTCCACGACT  
 25 601 CAAGGACTTA CCATAGAGAA TAACCTAGGG ACGCTAAAGT TCAATGAAAA  
 651 CAAAGCAGTG ACCTCAGGAG GCGCCTTAGA TTTAGGAGCC GCGTCTACAT  
 701 TCACTGCGAA CCATGAGTTG ATATTTTCAC AAAATAAGAC TTCTGGGAAT  
 751 GCTGCAAAATG GCGGAGCCAT AAATTGCTCA GGGGACCTTA CATTTACTGA  
 801 TAACACTTCT TTGTFACTTC AAGAAAATAG CACAATGCAG GATGGTGGAG  
 30 851 CTTTGTGTAG CACAGGAACC ATAAGCATTA CCGGTAGTGA TTCTATCAAT  
 901 GTGATAGGAA ATACTTCAGG ACAAAAAGGA GGAGCGATTT CTGCAGCTTC  
 951 TCTCAAGATT TTGGGAGGCG AGGGAGGCGC TCTCTTTTCT AATAACGTAG  
 1001 TGACTCATGC CACCCTCTA GGAGGTGCCA TTTTATCAA CACAGGAGGA  
 1051 TCCTTGCAGC TCTTCACTCA AGGAGGGGAT ATCGTATTCT AGGGGAATCA  
 35 1101 GGTCACATA ACAGCTCCAA ATGCTACCAC TAAGAGAAAT GTAATTCACC  
 1151 TCGAGAGCAC CGCGAAGTGG ACGGGAATTG CTGCAAGTCA AGGTAACGCT  
 1201 ATCTATTTCT ATGATCCCAT TACCACCAAC GATACGGGAG CAAGCGATAA  
 1251 CTTACGTATC AATGAGGTCA GTGCAAATCA AAAGCTCTCG GGATCTATAG  
 1301 TATTTTCTGG AGAGAGATTG TCGACAGCAG AAGCTATAGC TGAAAATCTT  
 40 1351 ACTTCGAGGA TCAACCAGCC TGTCACCTTA GTAGAGGGGA GCTTAGTACT  
 1401 TAAACAGGGA GTGACCTTGA TCACACAAGG ATTCTCGCAG GAGCCAGAA  
 1451 CCACGCTTCT TTTGGATCTG GGGACCTCAT TATAA

The PSORT algorithm predicts outer membrane (0.935).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 67A). The  
 45 recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure  
 67B) and for FACS analysis.

These experiments show that cp0018 is a surface-exposed and immunoaccessible protein, and that it  
 is a useful immunogen. These properties are not evident from the sequence alone.

#### Example 68

50 The following *C.pneumoniae* protein (PID 4376262) was expressed <SEQ ID 135; cp6262>:

1 MRKLRLAIV LIALSIILIA GGVLLTVAI PGLSSVISSP AGMGACALGC  
 51 VMLALGIDVL LKKREVPIVL ASVTTTPGTG SPRSGISISG ADSTIRSLPT  
 101 YLLDEGHPQS MRKLRLAIV LIVFSIILIA SGVLLTVAI PGLSSVISSP  
 151 AGMGACALGC VMLALGIDVL LKKREVPIVL ASVTTTPGTG SPRSGISISG  
 55 201 ADSTIRSLPT YPLDEGHPQS MRKLRLAIV LIVFSIILIA SGVLLTVAI  
 251 PGLSSIISP AEMGACALGC VMLALGIDVL LKKREVPIV PAPIPEEVVI

-108-

301 DDIDEESIRL QQAEAAALAR LPEEMSAFEG YIKVVESHLE NMKSLPYDGH  
 351 GLEEKTKHQI RVVRSILKAM VPEFLDIRRI FREERFFFLS ARKRLIDLAT  
 401 TLVERKILTE QLERNNLRKA PSYLYQDSIF KKIIDNFEKL AWKFMILSKS  
 451 ICRFTIIFEN HEHGVAKSLH HKNVLLLEKV IYRSLOKSYR DIGMSSAKMK  
 501 ILHGNPFFSL EDNKKTIMKE HAEMLESLSS YRKVFLALSD ENVVDTPSDP  
 551 KKWDLSGIPC RDALSEISRD EQWQKKAHLK HQESLYTQAR DRLTDQSSKE  
 601 NQKELEKAEQ EYISSWERVK KFEIERVQER IRAIQKLYPN ILEREBBTTG  
 651 QETVTPVQGG TTASSDLTDI LGRIEVSSRE DNQONQESCVK VLRSHVEVMS  
 701 WEVKQBYGPK KKEFQDQMG S LERFFTEHIE ELEVLOKDYS KHLSYFKKVN  
 751 NKKEVQYAKF RLKVLSDLEB GILAQTESAB SLLTQBELPI LATRGALSKA  
 801 VFKGSLCCAL ASKAKPYFEE DPRFQSDTQ LRLTLRLQE AKASLEEEIK  
 851 RFSNLENDIA KERRLLKESK QTFERAGLG V LREIAVESTY DLRLSLTNTWE  
 901 GTPESKQVYF SMYLNYYNEE KRRAKTRLVE MTQRYRDFKM ALEAMQFNEE  
 951 ALLQBELSIQ APSE\*

A predicted signal peptide is highlighted.

The cp6262 nucleotide sequence <SEQ ID 136> is:

1 ATGAGGAAAC TFCGTATTCT TCGATCGTT CTCATAGCTT TGAGCATTAT  
 51 TTTGATTGCA GGTGGTGTGG TATTGCTTAC TGTAGCGATC CCTGGATTAA  
 101 GTTCAGTCAT TTCTTCCCCG GCAGGATGG GTGCCGTGTC TTTGGGATGT  
 151 GTGATGCTTG CTTTAGGGAT CGATGTCTT CTGAAGAAAC GAGAAGTCCC  
 201 TATAGTTCTC GCATCTGTAA CTACGACACC AGGAAGTGGC AGCCCTAGAA  
 251 GTGGTATTTC TATTTAGGGA GCTGATAGCA CCATACGTTT TCTTCTACG  
 301 TATCTCTTGG ACGAGGGACA TCCACAATCC ATGAGGAAAC TFCGTATTCT  
 351 TCGCATCGTT CTCATAGTTT TAGCATTAT TTTGATTGCA AGTGGTGTGG  
 401 TATTGCTTAC TGTAGCGATC CCTGGATTAA GTTCAGTCAT TTCTTCCCCG  
 451 GCAGGATGG GTGCCGTGTC TTTGGGATGT GTGATGCTTG CTTTAGGGAT  
 501 CGATGTTCTT CTGAAGAAAC GAGAAGTCCC TATAGTTCTC GCATCTGTAA  
 551 CTACGACACC AGGAAGTGGC AGCCCTAGAA GTGGTATTTC TATTTAGGGA  
 601 GCTGATAGCA CCATACGTTT TCTTCTACG TATCCCTTGG ACGAGGGACA  
 651 TCCACAATCC ATGAGGAAAC TFCGTATTCT TCGATCGTT CTCATAGTTT  
 701 TAGCATTAT TTTGATTGCA AGTGGTGTGG TATTGCTTAC TGTAGCGATC  
 751 CCTGGATTAA GCTCGATCAT TTCTTCCCCA GCGGATGG GTGCTGTGTC  
 801 TTTGGGATGT GTGATGCTTG CTTTGGGGAT CGACGTCTT CTGAAGAAAC  
 851 GAGAAGTCCC TATAGTAGTT CCCGCACCTA TTCCTGAAGA AGTCGTCATA  
 901 GATGATATAG ATGAAGAGAG TATACGGCTG CAGCAGGAAG CTGAAGCCGC  
 951 TTTAGCAAGA CTTCTGAGG AGATGAGTGC ATTTGAAGGT TACATAAAG  
 1001 TTGTCGAGAG TCATTGGAG AACATGAAA GCCTGCCTTA TGATGGTCAT  
 1051 GGGCTAGAAG AGAAAACGAA ACATCAGATA AGAGTCGTCA GATCTTCTTT  
 1101 GAAGGCTATG GTTCCAGAA TTTTAGATAT CAGAAGAATT TTTGAAGAAG  
 1151 AAGAGTTCTT TTTTCTCTCA GCTCGCAAAC GACTTATAGA TTTAGCTACT  
 1201 ACTTTAGTAG AGAGAAAAAT TTTAACAGAG CAACCTGAGC GCAATAATTT  
 1251 AAGGAAAGCG TTTTCTTATT TATATCAGGA CTCAATTTTT AAAAAAATTA  
 1301 TTGATAACTT CGAGAAGTTA GCATGGAAT TTATGATTTT GAGTAAATCA  
 1351 ATTTGTCGAT TTACAATTAT TTTGAAAAT CATGAACATG GTGTAGCAAA  
 1401 GAGCCTGTTA CACAAGAATG CAGTGTTACT GGAGAAGGTA ATCTATAGGA  
 1451 GTTTGCAAAA AAGCTATAGA GATATAGGCA TGTCATCTGC AAAGATGAAA  
 1501 ATCTTGCACG GCAACCTTT TTTCTCTTTG GAAGATAATA AAAAGACGAT  
 1551 AATGAAAGAA CACGCAGAGA TGCTTGAAAG TCTCAGTAGC TATAGGAAGG  
 1601 TATTTTTPAGC TCTATCTGAT GAGAACGTTG TAGATACACC TAGCGATCCA  
 1651 AAGAAATGGG ATTTGTGAGG AATCCCCTGT AGGGACGCGT TGCTTGAGAT  
 1701 TTCTCGTGAT GAACAGTGGC AGAAGAAAGC ACATCTAAAG CATCAAGAGT  
 1751 CCCTCTATAC GCAAGCTAGG GATCGTTTAA CAGACCAGAG CTCTAAAGAA  
 1801 AATCAGAAAG AGTTAGAGAA AGCTGAACAA GAGTACATAT CTCTTTGGGA  
 1851 ACGGGTTAAA AAATTTGAGA TTGAGAGAGT ACAGGAGAGG ATACGGGCAA  
 1901 TTCAAAGCT TTATCTTAAT ATCCTCGAGA GAGAAGAAGA AACCACAGGT  
 1951 CAGGAGACTG TGACTCCAAC TGTTCAAGGG ACGACGGCTT CATCCGATTT  
 2001 AACAGATATT TTAGGAAGAA TAGAGGCTC CAGTAGGGAG GATAATCAGA  
 2051 ATCAAGAGTC TTGTGTAAAA GTCTTAAGAA GTCATGAGGT AGAAATGAGC  
 2101 TGGGAAAGTCA AACAAGAGTA TGGCCCTAAG AAAAAAGAA TTCAGGATCA  
 2151 AATGGGTTCT TTAGAGAGGT TTTTACAGA GCATATTGAA GAGTTAGAAG  
 2201 TATTACAGAA GGACTIONT AAACACTGT CTATTTTTAA AAAAGTAAAC  
 2251 AATAAGAAAG AGGTTCAATA TGCGAAGTTT AGGTTGAAGG TTTTAGAGTC  
 2301 AGATTTAGAA GGGATCTAG CTCAGACTGA GAGTGCTGAG AGTCTGTTAA  
 2351 CTCAAGAGA ACTTCCGATT CTTGCAACTC GGGGAGCCTT AGAGAAAGCT  
 2401 GTTTTCAAGG GGAGTCTATG TTGCGCGCTA GCAAGCAAAG CAAAACCTTA

-109-

2451 TTTTGAAGAG GATCCAGAT TCCAAGATTC TGATACGCAA TTGCGAGCTC  
 2501 TGACTCTAAG GTTACAGGAG GCTAAGGCAA GCCTGGAAGA AGAGATAAAG  
 2551 AGATTTTCAA ATCTTGAGAA CGATATTGCA GAGGAAAGAC GCCTTCTTAA  
 2601 AGAGAGCAAG CAGACGTTTCG AAAGAGCAGG TTTAGGGGTT CTCCGAGAAA  
 2651 TTGCAGTCGA GTCTACTTAT GATTTCGCTT CCTTAACAAA TACATGGGAA  
 2701 GGGACCCAG AGAGTGAGAA GGTCTATTTT AGCATGTATC TTAATTATTA  
 2751 CAACGAAGAG AACGTAGGG CTAAAACAAG ATTGGTTGAA ATGACACAGA  
 2801 GGTATAGAGA TTTTAAAATG GCCTTGGAAG CTATGCAGTT TAATGAAGAA  
 2851 GCCCTTTTGC AAGAGGAAC TCTATTCAA GCTCCAGTG AATAA

10 The PSORT algorithm predicts inner membrane (0.660).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 68A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 68B) and for FACS analysis.

These experiments show that cp6262 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 69

The following *C.pneumoniae* protein (PID 4376269) was expressed <SEQ ID 137; cp6269>:

1 MYQENLRLLLE RLLYNSVQKS YADRLFSYK TKMVHDTPLI PWEEKKEKCA  
 51 EAERKAFLEQQ KILLDYGKSI FWLNENDEIN LNDPWSWGLN TVRTRKVFQE  
 101 VDDSERWNHK VLIQKLEDDY EKLLLESSKE STEANKKLLS DLVDRLEDAK  
 151 TKFFFLKKQEE VETRVKDLRA RYGGTVDPKQ DTEAKKKVEL EASLETFLDS  
 201 IESELVQCLE DQDIYWKEQD VKDLARTQEL EEQDIEAKRE EAAEDLRSLN  
 251 ERLKRSKTM LDRKWHIENA EDSITWWT SQ IEMKDMKARL KILKEDITSV  
 301 LPEIDEIETC LSLEELPLLT TRELLTKSYL KFKICSETLL KMTSVFENNI  
 351 YVQBYEVQLQ NLGFKLQGIS QRFGKKQDDF ANLEBQVALQ KKRLRELTON  
 401 FVIQGFNFMK EDFKAAAKDL YIRSTAEQKM NFDVPCMELF RRYHHEVNKP  
 451 LLELMYNCAD SYRDAKKKLC SLRLDEKELL QKEIKKEEFY QKKQQRHADR  
 501 SRHTTYQKLR IAEELALELK KKI\*

The cp6269 nucleotide sequence <SEQ ID 138> is:

1 ATGTACCAGG AGAATCTAAG ATTGTTGGAA AGGCTTCTTT ATAATAGTGT  
 51 TCAAAAGAGC TATGCGGATC GGCTGTTTTT CTATGAAAAG ACAAAGATGG  
 101 TGCACGATAC TCCGCTGATT CCTTGGGAAG AGGATAAGGA AAAATGTGCT  
 151 GAAGCTGAGA AAGCTTTCTT AGAGCAACAG AAGATTCTCC TAGATTATGG  
 201 AAAATCTATC TTTTGGCTGA ATGAGAACGA TGAGATCAAT TTAAACGATC  
 251 CTTGGAGTTG GGGTCTTAAT ACGGTGAGGA CTAGGAAAGT ATTCCAAGAG  
 301 GTTGACGACA GTGAACGTTG GAATCATAAG GTACTCATTC AAAAAGTCTGA  
 351 GGACGATTAT GAGAACTTTC TAGAGGAAAG TTCAAAGAG TCTACTGAAG  
 401 CAAATAAGAA GCTTTTATCT GACTTAGTAG ATCGTCTTGA AGATGCTAAG  
 451 ACAAATTTT TCCTGAAGAA ACAGGAGGAG GTGGAGACTC GCGTTAAGGA  
 501 TCTTAGAGCT CGATATGGAG GCACAGTAGA TCCTAAGCAG GATACGGAAG  
 551 CTAAGAAGAA AGTCGAATTG GAGGCTAGCT TAGAAACCTT TTTAGATTCC  
 601 ATCGAATCAG AGCTAGTACA GTGTTTAGAA GATCAAGATA TATATTGGAA  
 651 AGAACAGGAT GTCAAAGATC TAGCACGTAC GCAAGAGCTC GAGGAACAAG  
 701 ATATTGAAGC GAAGAGGGAA GAAGCTGCCG AAGACCTAAG AAGTCTTAAT  
 751 GAGCGTTTAA AGAAGTCAAA AACTATGTTA GATAGGGCTA AATGGCATAT  
 801 TGAAAATGCT GAGGACAGTA TTACCTGGTG GACTAGTCAG ATAGAAATGA  
 851 AGGATATGAA AGCAAGACTG AAGATCTTAA AAGAAGATAT AACAAGTGTT  
 901 CTACCTGAAA TAGATGAGAT TGAAACGTGT TTAAGCTTAG AGGAGCTTCC  
 951 TTTGCTTACG ACCAGGGAAC TCTTAACTAA GTCCTACCTA AAGTTTAAGA  
 1001 TTTGTTTCGGA AACACTATTA AAAATGACTT CTGTGTTTGA GAACAATATC  
 1051 TATGTTTCAGG AGTACGAGGT TCAGCTGCAA AATCTAGGGT TTAAGTTACA  
 1101 AGGTATATCT CAGAGATTTCG GAAAGAAACA AGACGATTTT GCGAATCTAG  
 1151 AGGAACAGGT TGCTTTGCAA AAGAAACGAC TCAGAGAGCT CACTCAGAAT  
 1201 TTTGAAATAC AAGGATTCAA TTTTATGAAA GAAGATTTTA AGGCAGCCGC  
 1251 TAAAGATCTT TATATAAGAA GTACAGCTGA ACAAAGATG AACTTTGATG  
 1301 TGCCTTGCAT GGAGCTCTTC CGTAGGTATC ATGAGGAGGT CAACAAGCCG  
 1351 CTTCTTGAGT TGATGTACAA TTGTGCAGAC AGTTATAGAG ATGCTAAGAA

-110-

```

1401 AAAGCTTTGC TCTCTACGTC TTGATGAAAA AGAGTTATTA CAAAAAGAAA
1451 TCAAGAAAGA GGAATTTTAT CAAAAGAAAC AACAAAGGCA TGCAGATAGA
1501 TCACGTCATA CTACGTATCA AAAGCTACGA ATTGCTGAAG AGCTTGCTCT
1551 TGAGCTGAAG AAGAAATCT AA

```

5 The PSORT algorithm predicts cytoplasmic location (0.412).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 69A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 69B) and for FACS analysis.

These experiments show that cp6269 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 70

The following *C.pneumoniae* protein (PID 4376270) was expressed <SEQ ID 139; cp6270>:

```

1 MKIPLRFLLI SLVPTLSMSN LLGAATTEEL SASNSFDGTT STTSFSSKTS
51 SATDGTNYVF KDSVVIENVP KTGETQSTSC FKNDAAAGDL NFLGGGFSFT
101 FSNIDATTAS GAAIGSEAN KTVTLSGFSA LSFLKSPAST VTNGLGAINV
151 KGNLSLLDND KVLIQDNFST GDGGAINCAG SLKIANNKSL SFIGNSSSTR
201 GGAIHTKNLT LSSGGETLFQ GNTAPTAAGK GGAIAIADSG TLSISGDSGD
251 IIFEGNTIGA TGTVSHSAID LGTSAKITAL RAAQGHTIYF YDPITVTGST
301 SVADALNINS PDTGDNKEYT GTIVFSGEKL TEAEAKDEKN RTSKLLQNV
351 FKNGTVVLKG DVVLSANGFS QDANSKLIMD LGTSLVANTE SIELTNLEIN
401 IDSLRNGKKI KLSAATAQKD IRIDRPVULA ISDESPYQNG FLNEDHSYDG
451 ILELDAGKDI VISADSRSID AVQSPYGYQG KWTINWSTDD KKATVSWAKQ
501 SFNPTEAQEA PLVPNLLWGS FIDVRSFQNF IELGTEGAPY EKRFPWAGIS
551 NVLHRSGREN QRKFRHVSOG AVVGASTRMP GGDTLSLGFA QLFARDKDYF
601 MNTNFAKTYA GSLRLQHDAS LYSVVSILLG EGGLREILLP YVSKTLPCSF
651 YGQLSYGHTD HRMKTESLPP PPPTLSTDHT SWGGYVWAGE LGTRVAVENT
701 SGRGFFQEYT PFVKVQAVYA RQDSFVELGA ISRDFSDDLH YNLAIPGK
751 LEKRFAEQYY HVVAMYSPIV CRSNPKCTTT LLSNQGSWKT RGSNLRQAG
801 IVQASGFRSL GAAAEFLGNF GFWRGSSRS YNVDAGSKIK F*

```

30 A predicted signal peptide is highlighted.

The cp6270 nucleotide sequence <SEQ ID 140> is:

```

1 ATGAAGATTC CACTCCGCTT TTTATTGATA TCATTAGTAC CTACGCTTTC
51 TATGTCGAAT TTATTAGGAG CTGCTACTAC CGAAGAGTTA TCGGCTAGCA
101 ATAGCTTCGA TGGAACTACA TCAACAACAA GCTTTTCTAG TAAACATCA
151 TCGGCTACAG ATGGCACCAA TTATGTTTTT AAAGATTCTG TAGTTATAGA
201 AAATGTACCC AAAACAGGGG AAACCTAGTC TACTAGTTGT TTTAAAAATG
251 ACGCTGCAGC TGGAGATCTA AATTTCTTAG GAGGGGATT TCTTTTACA
301 TTTAGCAATA TCGATGCAAC CACGGCTTCT GGAGCTGCTA TTGGAAGTGA
351 AGCAGCTAAT AAGACAGTCA CGTTATCAGG ATTTTCGGCA CTTTCTTTTC
401 TTAAATCCCC AGCAAGTACA GTGACTAATG GATTGGGAGC TATCAATGTT
451 AAAGGGAATT TAAGCCTATT GGATAATGAT AAGGTATTGA TTCAGGACAA
501 TTTCTCAACA GGAGATGGCG GAGCAATTAA TTGTGCAGGC TCCTTGAAGA
551 TCGCAACAAA TAAGTCCCTT TCTTTTATTG GAAATAGTTC TTCAACACGT
601 GGCGGAGCGA TTCATACCAA AAACCTCACA CTATCTTCTG GTGGGGAAC
651 TCTATTTCAG GGAATACAG CGCCTACGGC TGCTGGTAAA GGAGGTGCTA
701 TCGCGATTGC AGACTCTGGC ACCCTATCCA TTTCTGGAGA CAGTGGCGAC
751 ATTATCTTTG AAGGCAATAC GATAGGAGCT ACAGGAACCG TCTCTCATAG
801 TGCTATTGAT TTAGGAATA GCGCTAAGAT AACTGCGTTA CGTGCTGCGC
851 AAGGACATAC GATATCTTT TATGATCCGA TTACTGTAAC AGGATCGACA
901 TCTGTTGCTG ATGCTCTCAA TATTAATAGC CCTGATACTG GAGATAACAA
951 AGAGTATACG GGAACCATAG TCTTTTCTGG AGAGAAGCTC ACGGAGGCAG
1001 AAGCTAAAGA TGAGAAGAAC CGCACTTCTA AATTACTTCA AAATGTTGCT
1051 TTTAAAAATG GGACTGTAGT TTTAAAAGGT GATGTCGTTT TAAGTGCAGAA
1101 CGGTTTCTCT CAGGATGCAA ACTCTAAGTT GATTATGGAT TTAGGGACGT
1151 CGTTGGTTGC AAACACCGAA AGTATCGAGT TAACGAATTT GGAAATTAAT
1201 ATAGACTCTC TCAGGAACGG GAAAAAGATA AAACCTAGTG CTGCCACAGC

```

-111-

1251 TCAGAAAGAT ATTTCGTATAG ATCGTCCTGT TGTACTGGCA ATTAGCGATG  
 1301 AGAGTTTTTA TCAAAATGGC TTTTGAATG AGGACCATTG CTATGATGGG  
 1351 ATTCTTGAGT TAGATGCTGG GAAAGACATC GTGATTTCTG CAGATTCTCG  
 1401 CAGTATAGAT GCTGTACAAT CTCCGTATGG CTATCAGGGA AAGTGGACGA  
 5 1451 TCAATFGGTC TACTGATGAT AAGAAAGCTA CGGTTTCTTG GCCGAAGCAG  
 1501 AGTTTTTAATC CCACTGCTGA GCAGGAGGCT CCGTTAGTTC CTAATCTTCT  
 1551 TTGGGGTTCT TTTATAGATG TTCGTTCCCT CCAGAATTTT ATAGAGCTAG  
 1601 GTACTGAAGG TGCTCCTTAC GAAAAGAGAT TTTGGGTTGC AGGCATTTCC  
 1651 AATGTTTTGC ATAGGAGCGG TCGTGAAAAT CAAAGGAAAT TCCGTCATGT  
 10 1701 GAGTGGAGGT GCTGTAGTAG GTGCTAGCAC GAGGATGCCG GGTGGTGATA  
 1751 CCTGTCTCTT GGGTTTTGCT CAGCTCTTTG CGCGTGACAA AGACTACTTTT  
 1801 ATGAATACCA ATTTTCGCAA GACCTACGCA GGATCTTTAC GTTTGCAGCA  
 1851 CGATCTTCC CTATACTCTG TGGTGAGTAT CCTTTTAGGA GAGGGAGGAC  
 1901 TCCGCGAGAT CCTGTGCGCT TATGTTTCCA AGACTCTGCC GTGCTCTTTC  
 15 1951 TATGGGCAGC TTAGCTACGG CCATACGGAT CATCGCATGA AGACCGAGTC  
 2001 TCTACCCCCC CCCCCCCGA CGCTCTCGAC GGATCATACT TCTTGGGGAG  
 2051 GATATGTCTG GGCTGGAGAG CTGGGAACTC GAGTTGCTGT TGAAAATACC  
 2101 AGCGGCAGAG GATTTTTCCA AGAGTACACT CCATTTGTAA AAGTCCAAGC  
 2151 TGTTTACGCT CGCCAAGATA GCTTTGTAGA ACTAGGAGCT ATCAGTCGTG  
 20 2201 ATTTTAGTGA TTCGCATCTT TATAACCTTG CGATTCCTCT TGGAATCAAG  
 2251 TTAGAGAAAC GGTTCGAGA GCAATATTAT CATGTTGTAG CGATGTATTC  
 2301 TCCAGATGTT TGTCGTAGTA ACCCCAAATG TACGACTACC CTACTTTCCA  
 2351 ACCAAGGGAG TTGGAAGACC AAAGGTTCCA ACTTAGCAAG ACAGGCTGGT  
 2401 ATTGTTCAAG CCTCAGGTTT TCGATCTTTG GGAGCTGCAG CAGAGCTTTT  
 25 2451 CGGGAACTTT GGCTTTGAAT GCGGGGATC TTCTCGTAGC TATAATGTAG  
 2501 ATGCCGGTAG CAAAATCAAA TTTTAG

The PSORT algorithm predicts outer membrane (0.92).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 70A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot and for  
 30 FACS analysis (Figure 70B).

The cp6270 protein was also identified in the 2D-PAGE experiment (Cpn0013).

These experiments show that cp6270 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 71

35 The following *C.pneumoniae* protein (PID 4376402) was expressed <SEQ ID 141; cp6402>:

1 MNVADLLSHL ETLLSSKIFQ DYGPNGLQVG DPQTPVKKIA VAVTADLETI  
 51 KQAVAAEANV LIVHGGIFWK GMPYPITGMI HKRIQLLIEH NIQLIAYHLP  
 101 LDAHPTLGNN WRVALDLNWH DLKPFSSSLP YLGVQGSFSP IDIDSFIDLL  
 151 SQYYQAPLKG SALGGPSRVS SAALISGGAY RELSSAATSQ VDCFTTGNFD  
 40 201 EPAWSTALES NINFLAFGHT ATEKVGPKSL AEHLKSEFPI STTFIDTANP  
 251 P\*

The cp6402 nucleotide sequence <SEQ ID 142> is:

1 ATGAATGTTG CGGATCTCCT TTCTCATCTT GAGACTCTTC TCTCATCAAA  
 51 AATAATTTAG GATTATGGAC CCAACGGACT TCAAGTTGGA GATCCCCAAA  
 45 101 CTCCGGTAAA GAAAATCGCT GTTGCAAGTTA CCGCAGATCT AGAAACCATA  
 151 AAACAAGCTG TTGCGGCCGA AGCAAACGTT CTCATGTGAC ACCACGGAAT  
 201 TTTTGGGAAA GGTATGCCCT ATCCTATTAC CGGCATGATC CATAAGCGCA  
 251 TCCAATTACT AATAGAACAC AATATCCAAC TCATTGCCTA CCACCTTCCT  
 301 TTGGATCTCT ACCCTACCTT AGGAATAAC TGGAGAGTTG CCTTGGATCT  
 50 351 AAATTGGCAT GACTTGAAGC CCTTTGGTTC TTCCCTCCCT TATTTAGGAG  
 401 TGCAAGGCTC TTTCTCTCCT ATCGATATAG ATTCTTTTCAT TGACCTGTGA  
 451 TCTCAATATT ACCAAGCTCC CCTAAAAGGA TCTGCCCTTG GCGGCCCTC  
 501 TAGAGTCTCC TCAGCAGCTC TGATCTCAGG AGGAGCTTAT AGAGAACTCT  
 551 CTTCCGGCAGC CACGTCCCAA GTCGATTGCT TCATCACAGG AAATTTTGAT  
 55 601 GAACCTGCAT GGTGACAGC TCTAGAAAGC AATATCAACT TCCTAGCATT  
 651 TGGACATACA GCCACAGAAA AAGTAGGTCC AAAATCTCTT GCAGAGCATC

-112-

701 TAAAAAGCGA ATTTCTATT TCCACAACCT TTATAGATAC GGCCAACCCC  
751 TTCTAA

The PSORT algorithm predicts cytoplasmic (0.158).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 71A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 71B) and for FACS analysis.

These experiments show that cp6402 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 72

10 The following *C.pneumoniae* protein (PID 4376520) was expressed <SEQ ID 143; cp6520>:

1 MKHYLSFSPS ADFFSKQGA I ETQVLFGERV LVKGSTCYAY SQLFHNELLW  
51 KPYPGHSFRS TLVPCTPEFH IHPNVSVVSV DAFLEDFWGP LPFGTLLHVN  
101 SQNTVIFPKD ILNHMNTIWG SGTPQCDPRH LRRLNYNFFA ELLIKDADLL  
151 LNFPVWVGGR SVHESLEKPG VDCSGFINIL YQAQGYNVPR NAADQYADCH  
15 WISSFENLPS GGLIFLYPKE EKRIHVMLK QDSSTLIHAS GGGKKVEYFI  
251 LEQDGKFLDS TYLFPRNNQR GRAFFGIPRK RKAFL\*

The cp6520 nucleotide sequence <SEQ ID 144> is:

1 ATGAAACACT ACCTATCATT TTCTCCTTCT GCTGATTTTT TCTCTAAACA  
51 GGGTGCTATT GAAACTCAAG TCCTTTTTGG AGAGCGCGTC TTAGTCAAAG  
201 GGAGCACCTG CTATGCATAT TCCCAATTAT TCCACAATGA GCTGTTATGG  
151 AAGCCCTATC CAGGTCATAG CTTTCGTTCT ACCCTAGTCC CCGCACTCC  
201 TGAATTTTCA ATCCATCCAA ATGTTTCTGT GGTTCCTGTG GATGCATTTT  
251 TAGATCCTTG GGGGATCCCT CTTCTTTTGG GAACTTTACT CCATGTGAAT  
301 TCTCAAAATA CCGTTATTTT CCCTAAGGAT ATTCTCAATC ATATGAACAC  
251 CATCTGGGGC TCCGGCACAC CTCATGCGA TCCTAGACAT CTACGTCGTC  
401 TAAATTATAA CTTCTTTGCT GAACTTTTAA TTAAAGACGC AGACCTTTTA  
451 CTGAACTTTC CCTATGTATG GGGAGGACGG TCTGTACACG AAAGTCTGGA  
501 AAAGCCGGGT GTTGATTTGT CCGGATTTAT CAATATCCTT TACCAGGCAC  
551 AGGGATACAA CGTCCCTAGA AACGCTGCAG ATCAATATGC GGATGTCAT  
301 TGGATCTCTA GCTTTGAGAA CCTTCCTTCT GGTGGGTAA TATTTCTTTA  
651 CCCTAAAGAA GAAAAGCGTA TTTCTCATGT TATGTTGAAA CAGGATAGTT  
701 CCACCCTCAT TCATGCTTCT GGTGGAGGGA AAAAAGTGA GTATTTTCAAT  
751 TTAGAACAAG ATGGGAAGTT TTTAGATTCT ACTTATCTAT TTTTTAGAAA  
801 TAATCAGAGG GGACGGCAT TTTTGGGAT CCCTAGAAAA AGAAAAGCCT  
351 TTCTGTAA

The PSORT algorithm predicts cytoplasmic (0.265).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 72A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 72B) and for FACS analysis.

40 These experiments show that cp6520 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 73

The following *C.pneumoniae* protein (PID 4376567) was expressed <SEQ ID 145; cp6567>:

1 MTSPPIPFQSS GDASFLAEQP QQLPSTSESQ LVTQLLTMMK HTQALSETVL  
451 QQQRDLRPTA SIILQVGGAP TGGAGAPFQP GPADDDHHPI PPPVVPAQIE  
101 TEITITIRSEL QLMRSTLQOS TKGARTGVLV VTAILMTISL LAIIIIILAV  
151 LGFTGVLPQV ALLMQGETNL IWAMVSGSII CFIALIGTLG LILTNNKNTPL

-113-

201 PAS\*

The cp6567 nucleotide sequence &lt;SEQ ID 146&gt; is:

1 ATGACCTCAC CGATCCCCTT TCAGTCTAGT GCGGATGCCT CTTTCCTTGC  
 5 51 CGAGCAGCCA CAGCAACTCC CGTCTACTTC TGAATCTCAG CTAGTAACTC  
 101 AATTGCTAAC CATGATGAAG CATACTCAAG CATTATCCGA AACGGTTCTT  
 151 CAACAACAAC GCGATCGATT ACCAACCGCA TCTATATATCC TTCAAGTAGG  
 201 AGGAGCTCCT ACAGGAGGAG CGGGTGCGCC TTTTCAACCA GGACCGGCAG  
 251 ATGATCATCA TCATCCCATA CCGCCGCTTG TTGTACCAGC TCAAATAGAA  
 301 ACAGAAATCA CCACTATAAG ATCCGAGTGA CAGCTCATGC GATCTACTCT  
 10 351 ACAACAAAGC ACAAAGGAG CTCGTACAGG AGTTCTAGTG GTTACTGCAA  
 401 TCTTAATGAC GATCTCCTTA TTGGCTATTA TTATCATAAT ACTAGCTGTG  
 451 CTTGGATTGA CGGGCGTCTT GCCTCAAGTA GCTTTATTGA TGCAGGGTGA  
 501 AACAAATCTG ATTTGGGCTA TGGTGAGCGG TTCTATTATT TGCTTTATTG  
 551 CGCTAATTGG AACTCTAGGA TTAATTTTAA CAAATAAGAA CACGCCTCTA  
 15 601 CCGGCTTCTT AA

The PSORT algorithm predicts inner membrane (0.694).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 73A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 73B) and for FACS analysis.

- 20 These experiments show that cp6567 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

**Example 74**The following *C.pneumoniae* protein (PID 4376576) was expressed <SEQ ID 147; cp6576>:

1 MLIMRNKVL QISILALIQT PLTLFSTKEKV KEGHVVDISI TIITEGENAS  
 25 51 NKHPLPKLKT RSGALFSQLD FDEDLRLAK BYDSVEPKVE FSEGKTNIAL  
 101 HLIAPKPSIRN IHISGNQVVP EHKILKTLQI YRNDLPEREK FLKGLDDLRT  
 151 YYLKRGYFAS SVDYSLEHNQ EKGHIDVLIK INEGPCGKIK QLTFSGISRS  
 201 EKSDIQEFIQ TKQHSSTTSW FTGAGLYHPD IVEQDSLAIT NYLHNNGYAD  
 251 AIVNSHYDLD DKGNNILYMD IDRGSRYTIG HVHIQGFVFL PKRLIEKQSQ  
 30 301 VGPNDLYCPD KIWDGAHKIK QTYAKYGYIN TNVDVLFIPH ATRPIVDVTY  
 351 EVSEGSFPYKV GLIKITGNTH TKSDVILHET SLFPDGTFFNR LKLEDTEQRL  
 401 RNTGYFQSVS VYTVRSQLDP MGNADQYRDI FVEVKETTTG NLGLFLGFSS  
 451 LDNLFGGIEL SESNFDLFGA RNIFSKGFRC LRGGGEHLFL KANFGDKVTD  
 501 YTLKWKPHF LNTPWILGIE LDKSINRALS KDYAVQTYGG NVSTTYILNE  
 35 551 HLKYGLFPRG SQTSLHEKRIK PLLGPNIDSN KGFVSAAGVN LNYDSVDSPR  
 601 TPTTGIRGGV TFEVSGLGST YHFTKLSLNS SIYRKLTRKG ILKIKGEAQF  
 651 IKPYSNPTAE GVPVSEPFLL GGETTVRGYK SFIIGPKYSA TEPQGLSSSL  
 701 LISEEFQYPL IROPNISAFV FLDSGFVGLQ EYKISLKLDR SSAGFGLRFD  
 751 VMNNVFMVLG FGWPFPRPTET LNKEKIDVSQ RFFFPALGGMF \*

- 40 A predicted signal peptide is highlighted.

The cp6576 nucleotide sequence &lt;SEQ ID 148&gt; is:

1 ATGCTCATCA TGCAGAAATA AGTTATCTTG CAAATATCTA TTCTAGCGTT  
 51 AATCCAAACC CCTTTAACTT TATTTTCTAC TGAATAAGTT AAAGAAGGCC  
 101 ATGTGGTGGT AGACTCTATC ACAATCATAA CGGAAGGAGA AAATGCTTCA  
 45 151 AATAAACATC CCTTACCCAA ATTAAGAGACC AGAAGTGGGG CTCTTTTTC  
 201 TCAATTAGAT TTTGATGAAG ACTTGAGAAT TCTAGCTAAA GAATACGACT  
 251 CTGTTGAGCC TAAAGTAGAA TTTTCTGAAG GGAAACTAA CATAGCCCTT  
 301 CACCTAATAG CTAAACCCTC AATTCGAAAT ATTCTATCTC CAGGAAATCA  
 351 AGTCGTTCCCT GAACATAAAA TTCTTAAAAA CCTACAAATT TACCGTAATG  
 50 401 ATCTCTTTGA ACGAGAAAAA TTTCTTAAGG GTCTTGATGA TCTAAGAACG  
 451 TATTATCTCA AGCGAGGATA TTTTCGCATCC AGTGTAGACT ACAGTCTGGA  
 501 ACACAATCAA GAAAAAGGTC ACATCGATGT TTTAATTAAA ATCAATGAAG  
 551 GTCCTTGCGG GAAAAATTAA CAGCTTACGT TCTCAGGAAT CTCTCGATCA  
 601 GAAAAATCAG ATATCCAAGA ATTTATTCAA ACCAAGCAGC ACTCTACAAC



-114-

5 651 TACAAGTTGG TTTACTGGAG CTGGACTCTA TCACCCAGAT ATTGTTGAAC  
 701 AAGATAGCTT GGCAATTACG AATTACCTAC ATAATAACGG GTACGCTGAT  
 751 GCTATAGTCA ACTCTCACTA TGACCTTGAC GACAAAGGGA ATATTCTTCT  
 801 TTACATGGAT ATTGATCGAG GGTCCGCGATA TACCTTAGGA CACGTCCATA  
 851 TCCAAGGGTT TGAGGTTTTG CCAAAACGCC TTATAGAAAA GCAATCCCAA  
 901 GTCGGCCCCA ATGATCTTTA TTGCCCCGAT AAAATATGGG ATGGGGCTCA  
 951 TAAGATCAAA CAAACTTATG CAAAGTATGG CTACATCAAT ACCAATGTAG  
 1001 ACGTTCCTCT CATCCCTCAC GCAACCCGCC CTATTTATGA TGTAACCTAT  
 1051 GAGGTAAGTG AAGGGTCTCC TTATAAGTT GGGTTAATTA AAATTACTGG  
 1101 GAATACCCAT ACAAATCTG ACGTTATTTT ACACGAAACC AGTCTCTTCC  
 1151 CAGGAGATAC ATTCAATCGC TTAAAGCTAG AAGATACTGA GCAACGTTTA  
 1201 AGAAATACAG GCTACTTCCA AAGCGTTAGT GTCTATACAG TTCGTTCTCA  
 1251 ACTTGATCCT ATGGGCAATG CGGATCAATA CCGAGATATT TTTGTAGAAG  
 1301 TCAAAGAAAC AACAAACAGGA AACTTAGGCT TATTCTTAGG ATTTAGTTCT  
 1351 CTTGACAATC TTTTGGAGG AATTGAACTA TCTGAAAGTA ATTTTGATCT  
 1401 ATTTGGAGCT AGAAATATAT TTTCTAAAGG TTTTCGTTGT CTAAGAGGCG  
 1451 GTGGAGAACA TCTATTCTTA AAAGCCAAC TCGGGGACAA AGTCACAGAC  
 1501 TAACCTTTGA AGTGGACCAA ACCTCATTTT CTAACACTC CTTGGATTTT  
 1551 AGGAATTGAA TTAGATAAAT CAATTAACAG AGCATTATCT AAAGATTATG  
 20 1601 CTGTCCAAAC CTATGGCGGG AACGTCAGCA CAACGTATAT CTTGAACGAA  
 1651 CACCTGAAAT ACGGTCTATT TTATCGAGGA AGTCAAACGA GTTTACATGA  
 1701 AAAACGTAAG TTCCTCCTAG GGCCAAATAT AGACAGCAAT AAAGGATTTG  
 1751 TCTCTGCTGC AGGTGTCAAC TTGAATTACG ATTCTGTAGA TAGTCTTAGA  
 25 1801 ACTCCAAC TAAGGATTTC CGGGGGGGTG ACTTTTGAGG TTTCTGGTTT  
 1851 GGGAGGAACT TATCATTTTA CAAACTCTC TTTAAACAGC TCTATCTATA  
 1901 GAAACTTAC GCGTAAAGGT ATTTTGAAAA TCAAAGGGGA AGCTCAATTT  
 1951 ATTAACCCCT ATAGCAATAC TACAGCTGAA GGAGTTCCTG TCAGTGAGCG  
 2001 CTTCTTCCTA GGTGGAGAGA CTACAGTTCG GGGATATAAA TCCTTTATTA  
 2051 TCGGTCCAAA ATACTCTGCT ACAGAACCTC AGGGAGGACT CTCTTCGCTC  
 30 2101 CTTATTTTCAG AAGAGTTTCA ATACCCTCTC ATCAGACAAC CTAATATTAG  
 2151 TGCCTTTGTA TTCTTAGACT CAGGTTTGT CCGTTTACAA GAGTATAAGA  
 2201 TTTCTGTTAA AGATCTACGT AGTAGTGTG GATTGGTCT GCGCTTCGAT  
 2251 GTAAATGAATA ATGTTCTCTGT TATGTTAGGA TTTGGTTGGC CCTTCCGTCC  
 2301 AACCAGAGCT TTGAATGGAG AAAAAATTGA TGTATCTCAG CGATTCTTCT  
 35 2351 TTGCTTTAGG GGGCATGTTT TAA

The PSORT algorithm predicts outer membrane (0.7658).

The protein was expressed in *E.coli* and purified as GST-fusion (Figure 74A), his-tag and his-tag/GST-fusion products. The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (Figure 74B) and for FACS analysis (Figure 74C).

40 The cp6576 protein was also identified in the 2D-PAGE experiment (Cpn0300).

These experiments show that cp6576 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 75

The following *C.pneumoniae* protein (PID 4376607) was expressed <SEQ ID 149; cp6607>:

45 1 MNKRQKDKLK ICVIISTLIL VGIFARAPRG DTFKTFKSE EAIIVSNQCN  
 51 EDMRKILCDA IEHADREIFL RIYNLSEPKI QQSLTRQAQA KNKVITYYQK  
 101 FKIPQILKQA SNVTLVEQPP AGRKLMHQKA LSIDKKDAWL GSANYTNLSL  
 151 RLDDNNLILGM HSSLECDLII TMTSGDFSIR DQGTGYFVLP QDRKIAIQAV  
 201 LEKIQTAKQT IQVAMFALTH SEIIQALHQA KQGIHVDII IDRSHSKLTF  
 50 251 KQLRQLNINK DFVSTNTAPC TLHHKFAVID NKTLLAGSIN WSKGRPSLND  
 301 ESLIILENLT KQONQKLRLI WKDLAKHSEH PTVDDDEEKEI IEKSLPVEEQ  
 351 EAA\*

A predicted signal peptide is highlighted.

The cp6607 nucleotide sequence <SEQ ID 150> is:

-115-

```

1  ATGAATAAAA GACAAAAAGA TAAATTAAAA ATCTGTGTGA TTATTAGCAC
51  GTTGATTTTA GTAGGAATTT TTGCAAGAGC TCCTCGTGGT GACACTTTTA
101 AGACTTTTTT AAAGTCTGAA GAAGCTATCA TCTACTCAA TCAATGCAAT
5   151 GAGGACATGC GTAAAATTCT ATGCGATGCT ATAGAACACG CTGATGAAGA
    201 GATCTTCCTA CGTATTTATA ACCTCTCAGA ACCCAAGATC CAACAGAGTT
    251 TAACTCGACA AGCTCAAGCA AAAAACAAAG TTACGATCTA CTATCAAAAA
    301 TTTAAAATTC CCCAAATCTT AAAGCAAGCC AGCAATGTAA CTTTAGTCGA
    351 GCAACCTCCA GCAGGGCGTA AACTGATGCA TCAAAAAGCT CTTTCCATAG
10  401 ATAAGAAAGA TGCTTGGCTA GGATCTGCGA ACTACACCAA TCTTTCTCTA
    451 CGTTTAGATA ATAATCTCAT TCTAGGAATG CATAGCTCGG AGCTCTGTGA
    501 TCTCATTATC ACAAATACCT CTGGAGACTT TTCTATAAAG GATCAAACAG
    551 GAAAGTATTT TGTTCCTCCT CAAGATCGTA AAATTGCAAT ACAAGCTGTA
    601 CTCGAAAAAA TCCAGACAGC TCAGAAAACC ATCCAAGTTG CTATGTTTGC
    651 TCTGACCCAC TCGGAGATTA TTCAAGCCTT ACATCAAGCA AAACAACGAG
15  701 GAATCCATGT AGATATTATC ATTGATAGAA GTCATAGCAA ACTTACTTTT
    751 AAGCAATTAC GACAATTAAA TATCAATAAA GACTTGTGTT CTATAAATAC
    801 CGCACCTGT ACTCTTCACC ATAAGTTTGC AGTTATAGAT AATAAACTC
    851 TACTTGCAGG ATCTATAAAT TGGTCTAAAG GAAGATTCTC CTTAAATGAT
    901 GAAAGCTTGA TCATACTGGA AAACCTGACC AAACAACAAA ATCAGAAACT
20  951 TCGAATGATT TGGAAAGATC TAGCTAAGCA TTCAGAACAT CCTACAGTAG
    1001 ACGATGAAGA AAAAGAAATT ATAGAAAAAA GTCTTCCAGT AGAAGAGCAA
    1051 GAAGCAGCGT GA

```

The PSORT algorithm predicts periplasmic (0.934).

The protein was expressed in *E.coli* and purified as a his-tagged product (Figure 75A) and also as a  
 25 GST-fusion. The GST-fusion protein was used to immunise mice, whose sera were used in a Western blot (Figure 75B) and for FACS analysis.

These experiments show that cp6607 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

#### Example 76

30 The following *C.pneumoniae* protein (PID 4376624) was expressed <SEQ ID 151; cp6624>:

```

1  MDAKMGYIFK VMRWIFCFVA CGITFGCTNS GFQANSRPC ILSMNRMIHD
51  CVERVVGNRLL ATAVLIKGSLL DPHAYEMVKG DKDKLAGSAV IFCNGLGLEH
101  TSLRLRKHLEN NPNSVKLGER LIARGAFVPL EEDGICDPHI WMDLSIWKEA
151  VIEITEVLIE KFPEWSAEFK ANSEELVCEM SILDSWAKQC LSTIPENLRY
35  201 LVSGHNAFSY FTRRYLATPE EVASGAWRSR CISPEGLSPE AQISVRDIMA
    251 VVDYINSHDV SVVFPEDTLN QDALKKIVSS LKKSLLVRLA QKPLYSDNVD
    301 DNYFSTFKHN VCLITEELGG VALECQR*

```

The cp6624 nucleotide sequence <SEQ ID 152> is:

```

1  ATGGATGCGA AAATGGGATA TATATTAAAA GTGATGCGTT GGATTTTCTG
40  51  TTTCGTGGCA TGTGGTATAA CTTTGGGATG TACCAATTCT GGGTTTCAGA
    101 ATGCAAATTC ACGTCCCTGT ATACTATCCA TGAATCGCAT GATTTCATGAT
    151 TGTGTTGAAA GAGTCGTGGG GAATAGGCTT GCTACCGCTG TTTTGATCAA
    201 AGGATCCTTA GACCCATCATG CGTATGAGAT GGTAAAGGG GATAAGGACA
    251 AGATTGCTGG AAGTGCCGTA ATTTTGTGTA ACGGCCTGGG TCTTGAGCAT
45  301 ACATTAAAGTT TCGGGAAGCA TTTAGAAAAT AATCCCAATA GTGTCAAGTT
    351 AGGGGAGCGG TTGATAGCGC GTGGGGCCTT TGTTCCTCTA GAAGAAGACG
    401 GTATTTCGGA TCCTCATATC TGGATGGATC TTTCTATTTC GAAGGAAGCT
    451 GTCATAGAAA TTACAGAAGT TCTCATTGAA AAGTTCCTTG AATGGTCTGC
    501 TGAATTAAAA GCAAATAGTG AGGAACTTGT TTGTGAAATG TCTATTTTAG
50  551 ATTCTTGGGC GAAACAATGC TTGAGCACAA TTCCTGAAAA TTTACGGTAT
    601 CTGTCTCAGG GTCATAATGC GTTCAGTTAC TTTACACGTC GCTATTTTAGC
    651 TACTCCTGAA GAAGTGGCTT CCGGAGCATG GAGGTCTCGT TGTATTCTC
    701 CTGAGGCTCT ATCTCCAGAA GCTCAAATCA GTGTTCTGTA TATTATGGCG
    751 GTTGTAGATT ATATTAATGA GCATGATGTC AGTGTGGTTT TCCCTGAGGA
55  801 TACTCTGAAC CAAGATGCGT TGAATAAAT TGTTCCTCTC CTGAAGAAAA
    851 GTCATTTAGT TCGTCTAGCT CAAAACCAT TGTATAGTGA TAATGTGGAC
    901 GACAAATATT TTAGCACCTT TAAACATAAT GTCTGCCTTA TCACAGAAGA

```

-116-

951 ATTAGGAGGG GTGGCTCTTG AATGTCAAAG ATGA

The PSORT algorithm predicts inner membrane (0.168).

The protein was expressed in *E.coli* and purified as a his-tag product (Figure 76A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 76B) and for FACS analysis.

The cp6624 protein was also identified in the 2D-PAGE experiment.

These experiments show that cp6624 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 77

10 The following *C.pneumoniae* protein (PID 4376728) was expressed <SEQ ID 153; cp6728>:

```

1  MKSSVSWLFF SSIPLFSSLS IVAAEVTLDS SNNSYDGSNG TTFTVFSTTD
51  AAAGTTYSLL SDVSPQNA GA LGIPLASGCF LEAGGDLTFQ GNQHALKFAP
101 INAGSSAGTV ASTSAADKNL LFNDFSRLSI ISCPSLLLSP TGQCALKSVG
151 NLSLTGNSQI IFTQNFSSDN GGVINTKNFL LSGTSQFASF SRNQAFSGKQ
15  201 GGVVYATGTI TIENSPGIVS FSQNLAKGSG GALYSTDNCS ITDNFQVIFD
251 GNSAWAAQA QGGAICCTTT DKTVTLTGNK NLSFTNTAL TYGGAISGLK
301 VSIAGGPTL FQSNISGSSA GQGGGGAINI ASAGELALSA TSGDITFNNN
351 QVTNGSTSTR NAINIIDTAK VTSIRAATGQ SIYFYDPITN PGTAASTDTL
401 NLNLADANSE IEYGAIVFS GEKLSPTKA LANVTSTIR QPAVLARGDL
20  451 VLRDGVTVTF KDLTQSPGSR ILMGGTTLS AKRANLSLNG LAVNLSSLDG
501 TNKAALKTEA ADKNISLSGT IALIDTEGSP YENHNLKSAS TYPLLELTTA
551 GANGTITLGA LSTLTQEP E THYGYQGNWQ LSWANATSSK IGSINWTRTG
601 YIPSPERKSN LPLNSLWGNF IDIRSIQLI BTKSSGEPFE RELWLSGIAN
651 FFYRDSMPTR HGFRHISGGY ALGITATTPA EDQLTFAFQC LFARDRNHIT
25  701 GKNHGDYGA SLYFHTEGL FDIANFLWGR ATRAPVWLSE ISQIPLSPD
751 AKFSYLHTDN HMKTYTDDNS IIKGSWRNDA FCADLGASLP FVISVPYLLK
801 EVEPFVKVQY IYAHQQDFYE RHAEGRAPNK SELINVEIPI GVTFERDSKS
851 EKGTYDLTLM YILDAYRRNP KCQTSLIASD ANWMAYGTNL ARQGFVRAA
901 NHFQVNPME IFQQFAFEVR SSSRNYNTNL GSKFCF*
```

30 The cp6728 nucleotide sequence <SEQ ID 154> is:

```

1  ATGAAGTCCT CTGTCTCTTG GTTGTCTTTT TCTTCAATCC CGCTCTTTTC
51  ATCGCTCTCT ATAGTCGCGG CAGAGGTGAC CTTAGATAGC AGCAATAATA
101 GCTATGATGG ATCTAACGGA ACTACCTTCA CGGTCTTTTC CACTACGGAC
151 GCTGCTGCAG GAACTACCTA TTCTTACTT TCCGACGTAT CCTTTCAAAA
35  201 TGCAGGGGCT TTAGGAATTC CCTTAGCCTC AGGATGCTTC CTAGAAGCGG
251 GCGGCGATCT TACTTTCCAA GGAAATCAAC ATGCACTGAA GTTTGCATTT
301 ATCAATGCGG GCTCTAGCGC TGGAATGTA GCCAGTACCT CAGCAGCAGA
351 TAAGAACTCT CTCTTTAATG ATTTTCTAG ACTCTCTATT ATCTCTTGTC
401 CCTCTCTTCT TCTCTCTCCT ACTGGACAAT GTGCTTTAAA ATCTGTGGGG
40  451 AATCTATCTC TAACTGGCAA TTCCCAAATT ATATTTACTC AGAACTTCTC
501 GTCAGATAAC GCGGTGTTA TCAATACGAA AAACCTCTTA TTATCAGGGA
551 CATCTCAGTT TCGGAGCTTT TCGAGAAACC AAGCCTTCAC AGGGAAGCAA
601 GCGGTGTAG TTTACGCTAC AGGAACTATA ACTATCGAGA ACAGCCCTGG
651 GATAGTTTCC TTCTCTCAAA ACCTAGCGAA AGGATCTGGC GGTGCTCTGT
45  701 ACAGCACTGA CAACTGTTTC ATTACAGATA ACTTTCAAGT GATCTTTGAC
751 GGCAATAGTG CTTGGGAAGC CGCTCAAGCT CAGGCGGGGG CTATTTGTTG
801 CACTACGACA GATAAAACAG TGACTCTTAC TGGGAACAAA AACCTCTCTT
851 TCACAAATAA TACAGCATTG ACATATGGCG GAGCCATCTC TGGACTCAAG
901 GTCAGTATTT CCGCTGGAGG TCCTACTCTA TTTCAAAGTA ATATCTCAGG
50  951 AAGTAGCGCC GGTACGGGAG GAGGAGGAGC GATCAATATA GCATCTGCTG
1001 GGGAACTCGC TCTCTCTGCT ACTTCTGGAG ATATTTACCTT CAATAACAAC
1051 CAAGTCACCA ACGGAAGCAC AAGTACAAGA AACGCAATAA ATATCATTTGA
1101 TACCGCTAAA GTCACATCGA TACGAGCTGC TACGGGGCAA TCTATCTATT
1151 TCTATGATCC CATCACAAAT CCAGGAACCG CAGCTTCTAC CGACACATTG
55  1201 AACTTAAACT TAGCAGATGC GAACAGTGAG ATCGAGTATG GGGGTGCGAT
1251 TGTCTTTTCT GGAGAAAAGC TTTCCCTTAC AGAAAAGCA ATCGCTGCAA
```

-117-

1301 ACGTCACCTC TACTATCCGA CAACCTGCAG TATTAGCGCG GGGAGATCTT  
 1351 GFACTTCGTG ATGGAGTCAC CGTAACTTTC AAGGATCTGA CTCAAAGTCC  
 1401 AGGATCCCGC ATCTTAATGG ATGGGGGGAC TACACTTAGT GCTAAAGAGG  
 1451 CAAATCTTTC GCTTAATGGC TTAGCAGTAA ATCTCTCCTC TTTAGATGGA  
 1501 ACCAACAAGG CAGCTTTAAA AACAGAAGCT GCAGATAAAA ATATCAGCCT  
 1551 ATCGGGAACG ATTGCGCTTA TTGACACGGA AGGGTCATTC TATGAGAATC  
 1601 ATAACTTAAA AAGTGCTAGT ACCTATCCTC TTCTTGAAC TACCACCGCA  
 1651 GGAGCCAACG GAACGATTAC TCTGGGAGCT CTTTCTACCC TGA CTCTTCA  
 1701 AGAACCTGAA ACCCACTACG GGTATCAAGG AAACCTGGCAG TTGTCTTGGG  
 1751 CAAATGCAAC ATCCTCAAAA ATAGGAAGCA TCAACTGGAC CCGTACAGGA  
 1801 TACATTCTTA GTCTTGAGAG AAAAAGTAAT CTCCCTCTAA ATAGCTTATG  
 1851 GGGAAACTTT ATAGATATAC GCTCGATCAA TCAGCTTATA GAAACCAAGT  
 1901 CCAGTGGGGA GCCTTTTGGG CGTGAGCTAT GGCTTTCAGG AATTGCGAAT  
 1951 TTCTTCTATA GAGATTCTAT GCCCACCCTC CATGGTTTCC GCCATATCAG  
 2001 CGGGGGTTAT GCACTAGGGA TCACAGCAAC AACTCCTGCC GAGGATCAGC  
 2051 TTACTTTTGC CTTCTGCCAG CTCTTTGCTA GAGATCGCAA TCATATTACA  
 2101 GGTAAGAACC ACGGAGATAC TTACGGTGCC TCTTTGTATT TCCACCATAC  
 2151 AGAAGGGCTC TTCGACATCG CCAATTTCCT CTGGGGAAAA GCAACCCGAG  
 2201 CTCCCTGGGT GCTCTCTGAG ATCTCCAGA TCATTCTTTT ATCGTTCGAT  
 2251 GCTAAATTCA GTTATCTCCA TACAGACAAC CACATGAAGA CATATATAC  
 2301 CGATAACTCT ATCATCAAGG GTTCTTGGAG AAACGATGCC TTCTGTGCAG  
 2351 ATCTTGGAGC TAGCTGCCT TTTGTATTAT CCGTTCCGTA TCTTCTGAAA  
 2401 GAAGTCGAAC CTTTGTGCAA AGTACAGTAT ATCTATGCGC ATCAGCAAGA  
 2451 CTCTACGAG CGTCATGCTG AAGGACGCGC TTTCAATAAA AGCGAGCTTA  
 2501 TCAACGTAGA GATTCTTATA GGCGTCACCT TCGAAAGAGA CTCAAATCA  
 2551 GAAAAGGGAA CTTACGATCT TACTCTTATG TATATACTCG ATGCTTACCG  
 2601 ACGCAATCCT AAATGTCAA CTTCCCTAAT AGCTAGCGAT GCTAACTGGA  
 2651 TGGCCTATGG TACCAACCTC GCACGACAAG GTTTTCTGT TCGTGTGCG  
 2701 AACCATTTC AAGTGAACCC CCACATGGAA ATCTTCGGTC AATTTCCTTT  
 2751 TGAAGTACGA AGTTCTTCAC GAAATTATAA TACAAACCTA GGCTCTAAGT  
 2801 TTTGTTTCTA G

The PSORT algorithm predicts inner membrane (0.187).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 77A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 77B) and for FACS analysis.

The cp6728 protein was also identified in the 2D-PAGE experiment.

These experiments show that cp6728 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

#### Example 78

The following *C.pneumoniae* protein (PID 4376847) was expressed <SEQ ID 155; cp6847>:

1 MFVMKKLVRL CVVLLSLLPN VLFSSDLLRE EGIKKMMDKL IEYHVDAQEV  
 51 STDILSRSL SYIQSFDPHK SYLSNQEAV FLQSPETKKR LLKNYKAGNF  
 101 AIYRNINQLI HESILRARQW RNEWVKNPKE LVLEASSYQI SKQPMQWSKS  
 151 LDEVKQRQRA LLLSYLSLHL AGASSRYEG KEEQLAALCL RQIENHENVY  
 201 LGINDHGVAM DRDEEAYQPH IRVVKALAHS LDAHTAYFSK DEALAMRIQL  
 251 EKGCMGIGVV LKEDIDGVVV REIIPGGPAA KSGDLQLGDI IYRVDGKDIE  
 301 HLSFRGVLDL LRGGHGSTVV LDIHRGESDH TIALRREKIL LEDRRVDVSY  
 351 BPGDGVIGK VTLHSFYEGE NQVSSEQDLR RAIQGLKEKN LLGLVLDIRE  
 401 NTGGFLSQAI KVSGLFMTNG VVVVSRYADG TMKCYRTVSP KKFYDGPLAI  
 451 LVSKSSASAA EIVAQTLQDY GVALVVGDEQ TYGKGTIQHQ TITGDASQDD  
 501 CFKVTVGKYY SPGKSTQLQ GVKSDILIPS LYAEDRLGER FLEHPLPADC  
 551 CDNVLDHPLT DLDTQTRPWF QKYLPNLQK QETLWREMLP QLTKNSEQRL  
 601 SENSNFQAF L SQIKSSEKTD LSYGSNDLQL EBSINILKDM ILLQQCRK\*

A predicted signal peptide is highlighted.

The cp6847 nucleotide sequence <SEQ ID 156> is:

-118-

1 ATGTTTCGTAA TGAAAAAACT TGTCCGTCTA TGCCTAGTTC TTCTTTCTTT  
 51 ACTTCCGAAT GTATTATTTT CTTCCGATCT TTTACGAGAA GAGGGCATCA  
 101 AAAAGATGAT GGACAAGCTG ATCGAGTATC ATGTCGATGC TCAAGAGGTT  
 151 TCTACGGATA TACTCTCGCG TTCTTTATCT AGTTACATT C AATCTTTTGA  
 5 201 TCCTCATAAA TCTTATCTTT CAAACCAAGA GGTTCGAGTT TTCTACAGT  
 251 CTCCGGAAC AAAGAAACGT CTCTTAAAGA ATTATAAGGC AGGCAACTTT  
 301 GCTATTATATC GCAACATCAA TCAATTAAT CATGAGAGTA TTCTTCGTGC  
 351 CAGGCAGTGG AGAAACGAAT GGGTTAAGAA TCCAAAAGAG CTTGTATTGG  
 401 AGGCATCCTC ATATCAGATA TCGAAGCAAC CTATGCAATG GAGCAAATCT  
 10 451 TTAGACGAAG TGAAGCAGAG ACAACGCGCT CTACTCCTTT CCTATCTTTC  
 501 TTTACATCTT GCTGGAGCTT CTTCTCTCTG TTATGAGGGT AAAGAAGAGC  
 551 AGCTTGCTGC TCTGTGTCTA CGTCAAATCG AGAACCATGA GAATGTATAT  
 601 TTAGGTATCA ACGATCATGG TGTTCGTATG GATCGGGATG AAGAAGCCTA  
 651 CCAATTCCAT ATCCGTGTTG TTAAAGCTTT AGCTCATAGC TTAGATGCAC  
 15 701 ATACGGCGTA TTTCACTAAG GACGAAGCGT TGGCGATGCG AATCCAACCTA  
 751 GAAAAAGGCA TGTGTGGAAT TGGTGTGTGT CTGAAGGAAG ATATTGATGG  
 801 AGTTGTGTGT AGAGAAATCA TTCTGGGGG ACCTGCGGCT AAATCTGGGG  
 851 ATCTTCAGCT TGGAGATATC ATCTATCGGG TGGATGGCAA GGATATCGAG  
 901 CATCTTTCTT TCCGCGGTGT TTTAGATTGT TTACGTGGAG GTCATGGCTC  
 20 951 TACTGTAGTC TTAGATATCC ATCGTGGGGA GAGCGATCAT ACGATCGCCT  
 1001 TGAGAAGGGA GAAAAATCCTT TTAGAAGACC GTCGTGTGGA TGTTCCTTAT  
 1051 GAGCCTTATG GAGATGGTGT GATTGGGAAA GTTACGTTAC ATTCTTTTTA  
 1101 TGAAGGAGAA AATCAGGTTT CTAGTGAACA AGATCTACGT CGAGCGATTC  
 1151 AGGGATTAAA GGAGAAGAAC CTTCTTGGAT TAGTTTTAGA TATCCGAGAA  
 25 1201 AATACGGGTG GATTTTATAT TCAAGCGATC AAAGTTTCTG GTTATTTAT  
 1251 GACCAATGGC GTTGTGGTTG TATCTCGCTA TGCTGATGGT ACCATGAAGT  
 1301 GCTACCGCAC AGTATCTCCT AAAAAATCT ATGATGGTCC TTGGCTATT  
 1351 TTAGTATCTA AAAGTTCCGC ATCAGCAGCG GAGATTGTAG CACAACTCT  
 1401 CCAAGATTAT GGAGTTGCTT TAGTTGTTGG AGATGAGCAG ACCTATGGGA  
 30 1451 AGGGAACGAT TCAGCATCAA ACAATTACTG GAGATGCCTC TCAGGACGAT  
 1501 TGTTTTAAAG TTACTGTAGG GAAATATTAT TCCCCTTCTG GGAAATCGAC  
 1551 TCAACTTCAG GGAGTAAAAT CCGATATTTT AATTCTTCT CTCTATGCTG  
 1601 AAGATCGTCT AGGAGAGCGT TTTCTAGAGC ATCCCTTACC TGCAGATTGC  
 1651 TGTGATAATG TACTTCACGA TCCTCTCACG GACTTGGATA CTCAAACACG  
 35 1701 TCCTTGGTTT CAAAAATACT ATCTTCCTAA TCTACAAAAG CAAGAGACTC  
 1751 TTTGGAGAGA GATGCTACCT CAGCTTACGA AAAACAGTGA GCAAAGGCTT  
 1801 TCTGAGAATT CGAATTTFCA GGCATTTTGT TCGCAGATAA AATCATCTGA  
 1851 AAAAACGGAC CTATCTATG GTTCCAATGA TTTACAATTG GAAGAGTCGA  
 1901 TAAACATTTT GAAGGACATG ATTTTATTAC AACAGTGTAG AAAATAA

40 The PSORT algorithm predicts periplasmic (0.932).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 78A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 78B) and for FACS analysis.

45 These experiments show that cp6847 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 79

The following *C.pneumoniae* protein (PID 4376969) was expressed <SEQ ID 157; cp6969>:

1 MRLFSLGTY LFFSLALSSC CGYSILNSPY HLSSLGKSL QERIFIAPIK  
 51 EDPHGLCSA LTYELSKRSF AISGRSSCAG YTLKVELLNG IDKNIGFTYA  
 101 PNKLGDKTHR HFIVSNEGR LSLAKVQLIN NDTQEVLIQ CVARESVDFFD  
 151 FEPDLGTANA HEFALGQFEM HSEAIKSARR ILSIRLAETI AQQVYVDLP\*

A predicted signal peptide is highlighted.

The cp6969 nucleotide sequence <SEQ ID 158> is:

55 1 ATGAGATTGT TTCTTTAGG CACGATTTAT CTTTTTTTTT CTCTAGCACT  
 51 TTCGTATGCT TGTGGTFACT CTATTTTAAA CAGCCCGTAT CACTTATCGT  
 101 CTTTAGGTAA GTCTTTATTA CAGGAAAGAA TTTTCATTGC TCCCATAAAA

-119-

```

151 GAAGATCCTC ATGGTCAGCT CTGCTCAGCT CTAACCTATG AGCTTAGTAA
201 GCGTTCTTTT GCTATCTCTG GAAGGAGTTC TTGCGCAGGC TATACTCTTA
251 AAGTAGAGCT TCTGAATGGT ATTGACAAGA ATATAGGTTT TACGTATGCC
301 CCAAATAAAC TCGGAGATAA GACTCACAGG CATTTTATAG TCTCTAATGA
5 351 AGGCAGACTA TCACTATCTG CAAAAGTACA GCTTATCAAT AATGACACTC
401 AAGAAGTCCT TATAGACCAA TGTGTTGCTC GAGAGTCTGT AGACTTTGAC
451 TTTGAGCCTG ACTTAGGAAC AGCAAACGCT CATGAATTG CTTTAGGCCA
501 ATTTGAAATG CATAGTGAAG CCATAAAAG TGCTCGCCGT ATACTATCTA
551 TACGCCTAGC CGAGACGATT GCTCAACAGG TATACTATGA CCTTTTGTGA

```

10 The PSORT algorithm predicts inner membrane (0.126).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 79A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 79B) and for FACS analysis.

These experiments show that cp6969 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 80

The following *C.pneumoniae* protein (PID 4377109) was expressed <SEQ ID 159; cp7109>:

```

1 MKKTCCQNYR SIGVVFSVVL FVLTTQTLFA GHFIDIGTSG LYSWARGVSG
51 DGRVVVGYEG GNAFKYVDGE KFLLEGLVPR SEALVFKASY DGSVIIGISD
20 101 QDPSCRAVKW VNGALVDLGI FSEGMSQFAE GVSSDGKTIV GCLYSDDTET
151 NFAVKWDETG MVVLEPNLPED RHSCAWDASE DGSVIVGDAM GSEELAKAVY
201 WKDGEQHLLS NIPGAKRSSA HAVSKDGSFI VGEFISEENE VHAFVYHNGV
251 IKDIGTLGGD YSVATGVSRD GKVIVGHSTR TDGEYRAFKY VDGRMIDLGT
301 LGGSASFAGF VSDDGKTIVG KFETELGECH AFIYLLDD*

```

25 A predicted signal peptide is highlighted.

The cp7109 nucleotide sequence <SEQ ID 160> is:

```

1 ATGAAAAAGA CATGTTGCCA AAATTACAGA TCGATAGGCG TTGTGTTCTC
51 TGTGGTACTT TTCGTTCTTA CAACACAGAC GCTGTTTGCA GGACATTTTA
30 101 TTGATATTGG AACTTCTGGA TTATATTCTT GGGCTCGAGG TGTATCTGGA
151 GATGGCCGCG TTGTCGTAGG TTATGAAGGT GGCAATGCAT TTAAATATGT
201 TGATGGTGAG AAATTCTGTG TAGAAGGTTT GGTCCCGAGA TCCGAGGCCT
251 TGGTATTTAA AGCTTCTTAT GATGGCTCTG TAAATTATAGG AATCTCGGAT
301 CAAGATCCGT CTTGCCGCGC TGTGAAGTGG GTAAACGGTG CACTTGTGTA
351 TCTTGAATA TTTTCTGAGG GAATGCAATC TTTTGAGAG GGTGTTTCCA
35 401 GTGATGGAAA GACGATTGTA GGGTGCCTAT ATAGTGATGA TACAGAGACA
451 AACTTTGCTG TGAAGTGGGA TGAAACAGGA ATGGTTGTTT TCCCTAACTT
501 ACCAGAAGAT CGACATCTTT GCGCTTGGGA TGCCCTCTGAA GATGGCTCTG
551 TGATTGTAGG GGACGCCATG GGTAGCGAGG AAATTGCCAA GGCAGTGATC
601 TGGAAAGGACG GTGAACAACA TCTGCTTTCT AATATCCCAG GAGCTAAAAG
40 651 ATCGTCAGCA CATGCAGTTT CTAAAGATGG ATCTTTTATC GTAGGCGAGT
701 TCATCAGTGA AGAAAATGAA GTTCATGCCT TTGTTTATCA CAACGGTGTT
751 ATCAAAGATA TCGGGACTTT AGGAGGAGAT TACTCTGTAG CAACTGGAGT
801 TTCTAGGGAT GGTAAAGTCA TCGTGGGTCA TTCTACAAGA ACAGATGGTG
851 AATACCGTGC ATTTAAATAT GTGGATGGAA GAATGATAGA TTTGGGGACT
45 901 TTAGGAGGTT CAGCATCTTT TGCTTTTGGT GTTTCTGACG ATGGCAAAC
951 AATCGTAGGA AAATTGAAA CAGAGCTAGG AGAATGTCAT GCCTTTATCT
1001 ACCTTGATGA TTAG

```

The PSORT algorithm predicts outer membrane (0.887).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 80A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 80B) and for FACS analysis.

-120-

These experiments show that cp7109 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 81

The following *C.pneumoniae* protein (PID 4377110) was expressed <SEQ ID 161; cp7110>:

```

5      1  MAAIKQILRS  MLSQSSLWMV  LFSLYLSGY  CYVITDKPED  DFHSSSAVKW
      51  DHWGKTTLRS  LSNKKASAKA  VSGTGATTVG  FIKDTWSRTY  AVRWNWVGTK
     101  ELPTSSWVKK  SKATGISSDG  SIIAGIVENE  LSQSFVATWK  NNEMYLLPST
     151  WAVQSKAYGI  SSDGSVIVGS  AKDAWSRTFA  VKWTGHEAQV  LPVGWAVKSV
     201  ANSVSANGSI  IVGSVDASG  ILYAVKWEGN  TITHLGLTGG  YSALAKAVSN
    10  251  NGKVIIVGRSE  TYYGEVHAF  HKNGVMSDLG  TLGGSYSAAK  GVSATGKIVIV
     301  GMSTTANGKL  HAFKYVGGRM  IDLGEYSWKE  ACANAVSIDG  EIIIVGVQSE*
  
```

A predicted signal peptide is highlighted.

The cp7110 nucleotide sequence <SEQ ID 162> is:

```

15      1  ATGGCAGCTA  TAAAACAAAT  TTTACGTTCT  ATGCTATCTC  AGAGTAGCTT
      51  ATGGATGGTC  CTATTTTCAT  TATATTCCTC  ATCTGCTTAT  TGCTATGTAA
     101  TTACAGACAA  ACCAGAAGAT  GACTTCCATT  CTTTATCCGC  AGTAAAATGG
     151  GATCATTGGG  GAAAGACAAC  TCTCTCAAGA  TTATCAAATA  AAAAAGCCTC
     201  TGCAAAAGCT  GTTTCAGGAA  CTGGTGCTAC  AACTGTCGGC  TTTATAAAG
     251  ACACCTGGTC  TCGAACATAC  GCAGTAAGAT  GGAATTATTG  GGGGACCAAA
     301  GAACTCCCTA  CCAGCTCATG  GGTAAGAAAA  TCAAAAGCAA  CAGGAATCTC
     351  CTCTGATGGG  TCTATAATCG  CGGGGATTGT  CGAGAATGAG  CTTTCTCAAA
     401  GTTTCGCAGT  CACATGGAAA  AACAAAGAAA  TGTATTGCT  CCCTTCCACA
     451  TGGGCAGTGC  AATCTAAAGC  GTATGGAATT  TCTTCTGATG  GCTCTGTTAT
     501  TGTAGGGAGT  GCTAAGGATG  CTTGGTCGCG  AACTTTCGCT  GTGAAGTGGA
    25  551  CGGGACACGA  GGCTCAGGTG  TTACCAGTAG  GCTGGGCTGT  CAAATCTGTA
     601  CGGAATTCTG  TATCTGCCAA  TGGATCTATA  ATTGTAGGGT  CTGTACAAGA
     651  CGCCTCTGGA  ATTCTTTATG  CTGTAAAGTG  GGAAGGGAAC  ACTATTACAC
     701  ATCTAGGAAC  TTTAGGAGGC  TATCTGCCCA  TTGCAAAAGC  TGTATCCAAT
     751  AATGGCAAGG  TCATTGTAGG  GAGATCCGAA  ACATATTATG  GAGAGGTCCA
    30  801  TGCTTTCTGT  CATAAGAATG  GCGTCATGTC  AGACCTCGGC  ACCCTCGGAG
     851  GATCTTATTC  TGCAGCTAAG  GGAGTCTCTG  CAACTGGAAA  AGTTATTGTC
     901  GGTAGTGCCA  CAACAGCAAA  TGGGAAATTG  CATGCCTTTA  AATATGTCGG
     951  TGGAAGAATG  ATCGACTTAG  GAGAGTATAG  CTGGAAGAG  GCCTGTGCAA
    1001  ACGCTGTTTC  TATTGATGGA  GAAATTATTG  TTGGAGTCCA  ATCAGAATAA
  
```

35 The PSORT algorithm predicts outer membrane (0.827).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 81A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 81B) and for FACS analysis.

These experiments show that cp7110 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Figure 191 shows a schematic representation of the structural relationships between of cp7105, cp7106, cp7107, cp7108, cp7109 and cp7110, each of which is identified herein. These six proteins may be grouped in a new family of related outer membrane-associated proteins. These proteins have a repeat structure in common (*cf.* the pmp family).

### 45 Example 82

The following *C.pneumoniae* protein (PID 4377127) was expressed <SEQ ID 163; cp7127>:

```

1  MVFFRNSLLH  LVALSGMLCC  SSGVALTIAE  KMASLEHSGR  GADDYEGMAS
  
```

-121-

51 FNANMREYSL QLSKLYBEAR KLRASGTEDE ALWKDLIRRI GEVRGYLREI  
 101 EELWAAEIRE KGGNLEDYAL WNHPEPTTIYN LVTDYGTEDS IYLIPOEIGA  
 151 IKIATLSKFV VPKESEFEDCL TQILSRLGIG VRQVNSWIKE LYMMRKEGCS  
 201 VAGVFSSRKD LEALPETAYI GFVLNSNVDA HTNQHVLLKF INPETHVDV  
 5 251 IAGRUVIFGS AGEVGELLKI YNFVQSEISIR QEYRVIPLTK IDPGEMISIL  
 301 NAAFREDLTK DVSEESLGLR VVPLQYQGRS LFLSGTAALV QQALTLIREL  
 351 EREGIENPTDK TVFWYNVKHS DPQELAALLS QVHDFVSGEN KASVGAADGC  
 401 GSQNLASIQI DTTVSSSAKD GSVKYGNFIA DSKTGTLMV VEKEVLPRIQ  
 451 MLLKKLDVPR KMOVRIEVLFF ERKLAHEQKS GLNLLRLGEE VCKKGCSPSV  
 10 501 SWAGGTGILE FLFKGSTGSS IVPGYDLAYQ FLMAQEDVRI NASPSVVTMN  
 551 QTPARIAVVD EMSIAVSSDK DKAQYNRAQY GIMIKMLPVI NVGEEDGKSY  
 601 ITLETDTITFD TTGNHDDRP DVTRRNITNK VRIADGETVI IGGLRCQMS  
 651 DSHDGIPFLG DIPGIGKLF MSSTSDSLTE MFVFITPKIL ENPVEQQRK  
 701 EEALLSSRPG EREYYQALA ASEAAARAAH KKLEMPASG VLSQVERQE  
 15 751 YDGC\*

A predicted signal peptide is highlighted.

The cp7127 nucleotide sequence <SEQ ID 164> is:

1 ATGGTTTTTT TCCGTAATTC TTTACTGCAT TTAGTTGCCC TATCCGGAAT  
 51 GCTCTGTTGT TCTTCTGGAG TGGCTTTAAC GATAGCCGAG AAGATGGCTT  
 20 101 CTTTAGAGCA CTCGGGGAGA GGAGCAGACG ATTATGAGGG GATGGCTTCG  
 151 TTTAATGCCA ATATGAGGGA GTATAGCCTT CAGCTGAGCA AGTTGTATGA  
 201 GGAAGCACGA AAGCTACGCG CTTCTGGAAC TGAGGATGAA GCTCTGTGGA  
 251 AGGACTTAAT TCGACGGATT GGTGAGGTGC GAGGCTATCT TCGAGAGATC  
 301 GAGGAGCTTT GGGCTGCAGA AATTCTGTGAG AAAGGGGGCA ATCTCGAGGA  
 25 351 CTACGCCCTC TGGAAATCACC CAGAGACTAC GATTTACAAT CTTGTTACCG  
 401 ATTACGGAAC CGAAGACTCT ATTTATTTGA TTCTCAAGA AATCGGAGCG  
 451 ATTAATAATCG CAACCTTATC GAAATTTGTA GTTCTAAAG AGTCTTTTCA  
 501 AGACTGTCTC ACTCAGATCC TATCTCGCTT AGGTATTGGC GTGCGTCAGG  
 551 TCAATTCTTG GATTAAGGAA CTTTATATGA TCGTAAGGA GGGCTGCAGT  
 30 601 GTTGCTGGAG TTTTTCCTC CAGAAAAGAT TTAGAGGCGC TCCAGAAAC  
 651 AGCCTATATT GGTTTTGTAT TGAATTCGAA CGTAGATGCG CATACCAATC  
 701 AACATGTCTT AAAAAAGTTC ATTAACCTG AAACAACGCA TGATAGTGTG  
 751 ATTGACAGGAC GTGTGTGGAT TTTTGGTTCT GCGGGGGAAG TCGGCGAGCT  
 801 TCTGAAGATT TATAATTTTG TGCAGTCGGA GAGCATACGT CAAGAGTATC  
 35 851 GGGTGATTCC CTTAACTAAG ATCGATCCAG GGGAGATGAT TTCCATTCTC  
 901 AACGCAGCAT TTCGTGAGGA TCTGACTAAA GATGTTAGTG AAGAATCTTT  
 951 AGGCCTTCGT GTAGTTCCCT TACAGTATCA AGGCGCTTCG TTGTTTAA  
 1001 GTGGAACCGC GCGCTTAGTG CAGCAAGCGC TGAATCTCAT TCGAGAGCTT  
 1051 GAAGAAGGGA TTGAGAACC TACGATAAAA ACAGTATTTT GGTATAACGT  
 40 1101 CAAGCACTCC GATCCCCAAG AGTTGGCGGC ATTGCTTTCC CAAGTCCATG  
 1151 ATGCTCTCTC TGGCGAGAAT AAGGCGAGTG TCGGAGCTGC AGATGGATGT  
 1201 GGGTCGCAAT TAAATGCCTC GATCCAAATT GATACTACAG TAAGTTCTTC  
 1251 TGCGAAAGAT GGCTCAGTGA AGTACGGAAA CTTTATCGCG GATTCTAAGA  
 1301 CAGGAATCTT GATTATGGTG GTTGAGAAAG AAGTTCTTCC ACGTATTTCAG  
 45 1351 ATGCTACTTA AGAACTAGA TGTCCCTAAA AAGATGGTCC GTATCGAGGT  
 1401 GCTGTTATTT GAAAGAAAAT TGGCACATGA GCAGAAATCT GGGTTAAATC  
 1451 TTCTACGTCT TGGTGAGGAA GTTTGTAAAA AAGGGTGCAG TCCTTCTGTG  
 1501 TCTTGGGCCG GGGGTACTGG CATACTAGAA TTTTATTTA AAGGAAGTAC  
 1551 GGGATCTTCG ATAGTTCCCT GTTATGATCT CGCCTATCAA TTTTAAATGG  
 50 1601 CTCAAGAGGA CGTTCGGATT AATGCGAGTC CTTCTGTAGT TACTATGAAC  
 1651 CAAACCCAG CACGGATTGC TGTGTTGAT GAAATGTCAA TAGCGGTGTC  
 1701 TTCAGATAAA GATAAGCGC AATACAATCG TCGCAGTAC GGTATCATGA  
 1751 TAAAAATGCT CCCCCTAATT AATGTGGGAG AGGAAGACGG AAAAAGTTAC  
 1801 ATTACTTTAG AGACAGACAT CACCTTTGAT ACTACGGGA AAAATCATGA  
 55 1851 TGATCGTCTT GATGTTACAA GCGCTAATAT TACTAATAAG GTGCGATTG  
 1901 CTGACGAGGA GACTGTGATT ATTGGAGGTT TGGCTTGCAA ACAGATGTCA  
 1951 GATTCTCATG ATGGCATTCC TTTCTTTGGA GACATTCCTG GTATAGGGAA  
 2001 GTTATTTGGA ATGAGTTCCA CATCAGACAG TCTCACGGAG ATGTTTGTAT  
 2051 TTATCACTCC GAAGATCCTA GAAAACTCTG TAGAGCAACA AGAACGTAAA  
 60 2101 GAAGAAGCTT TACTCTCTTC GCGCCCTGGA GAGAGAGAAG AATACTATCA  
 2151 GGCTTTAGCA GCTAGTGAGG CTGCAGCACG AGCAGCTCAT AAAAAATTAG  
 2201 AGATGTTCCT GGCATCAGGA GTATCTTTAT CTCAGGTAGA GAGGCAAGAA  
 2251 TACGATGGCT GCTAG

The PSORT algorithm predicts periplasmic (0.920).



The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 82A) and also in his-tagged form. The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (Figure 82B) and for FACS analysis.

These experiments show that cp7127 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 83

The following *C.pneumoniae* protein (PID 4377133) was expressed <SEQ ID 165; cp7133>:

```

1  MQPFIFTLLC LTSLVSLVAF DAANARKRCA CAQTIERGEN FFSIKRSACA
51  EIEYQEKSRH ASAIERISKD KGKVTPKQIA KVATKKKQRY RLLQVPFSRP
101 PNNRYNLYA LLSEPPPECYS DTASWYAIFI RLLRRAYVDI GNVPPGSEYA
151 IANALISNKQ EILERGAQLG PDVIETLTLP EEQAEIFYKM LKGSNSQSL
201 LNFLHYREKS LGHCKLNLIF MDPLLEAVL DHPDAYRETS LLRDGIWEAV
251 KRQEHAIQEH GQAAALELFK TRTDFRLELR DKMQLLSRY DLLPLLNKKM
301 FDYTLGSAGD YLFLVDPDTK AISRCRCPSK SIKL

```

A predicted signal peptide is highlighted.

The cp7133 nucleotide sequence <SEQ ID 166> is:

```

1  ATGCAACCTT TTATCTTTAC TTTACTGTGC TTGACATCTT TGGTTTCTTT
51  AGTCGCCTTT GATGCTGCGA ATGCTCGTAA ACGTTGTGCC TGTGCTCAAA
101 CTATAGAACG TGGAGAGAAC TTCTTTTCCA TAAAACGCTC TGCTTGTGCT
151 GAAATCGAAT ATCAAGAAAA ATCTCGCCAC GCCTCAGCAA TTGAAAGAAT
201 CTCAAAGAT AAAGGCAAG TCACTCCAAA GCAGATTGCG AAAGTAGCTA
251 CTAAGAAAAA GCAAAGATAC CGTTTATTGC AGGTTCTTTT TTCAAGGCCT
301 CCGAATAACT CAAGGTATAA CCTCTATGCT TTGCTTAGTG AACCTCCCGA
351 ATGCTATAGC GATACAGCAT CATGGTATGC TATTTTATT TCGTTACTTC
401 GACGTGCTTA TGTAGACACG GGAAATGTAC CTCCTGGATC TGAGTATGCC
451 ATCGCTAATG CTTTGATAAG TAACAAACAA GAGATTTTAG AGAGGGGAGC
501 GCAGCTTGGG CCCGATGTTA TTGAAACTCT AACATTGCCT GAGGAACAAG
551 CCGAGATTTT TTATAAAATG CTCAAAGGGT CGTCAAACCT TCAGTCGCTA
601 CTGAATTTTC TGCATTATGA AGAGAAAAGC TTAGGCCACT GTAAGCTAAA
651 TCTGATCTTC ATGGATCCCC TACTGTTAGA AGCTGTTCTA GATCATCCCG
701 ATGCTTATAG GGAACGTCG CTCTGCGCG ATGGCATTG GGAAGCGGTG
751 AAGCGTCAAG AACATGCCAT CCAAGAACAT GGCCAGGCAG CTGCTTTTGA
801 GCTTTTTAAA ACACGCACCG ACTTCGCCT GGAGCTGCGA GATAAGATGC
851 AGTTACTTCT AAGTCGATAC GATTGCTCC CCTTATTAAA TAAAAAATG
901 TTCGACTACA CCTTAGGAAG TGCCGAGAT TACTTATTTT TGGTAGACCC
951 AGATACTAAG GCAATTTCTC GATGTCGCTG CCCTTCAAAG AGTATTAAAT
1001 TATAA

```

The PSORT algorithm predicts outer membrane (0.92).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 83A) and also in his-tagged form. The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (Figure 83B) and for FACS analysis.

These experiments show that cp7133 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 84

The following *C.pneumoniae* protein (PID 4377222) was expressed <SEQ ID 167; cp7222>:

```

1  MNRRDMVITA VVVNAILLVA LFVTSKRIGV KDYDEGFRNF ASSKVTOAVV
51  SEEKVIEKPV VAEVPSRPIA KETLAAQFIE SKPVIVTTPP VPVVSSETPEV

```

-123-

101 PTVAVPPQPV RETVKEEQAP YATVVVKGD FLERIANRH TTVAKLMQIN  
 151 DLTTTQLKIG QVIKVPSTQD VSNEKTPQTQ TANPENYYIV QEGDSPWTIA  
 201 LRNHIRLDDL LKMNDLDEYK ARRLKPGDQL RIR\*

A predicted signal peptide is highlighted.

5 The cp7222 nucleotide sequence <SEQ ID 168> is:

1 ATGAATCGTA GAGACATGGT AATAACAGCT GTCGTAGTGA ATGCTATATT  
 51 GCTTGTGGCT CTTTTCGTCA CATCAAAGCG TATTGGCGTC AAGGACTATG  
 101 ACGAGGGATT CCGTAATTTT GCTTCTAGCA AGGTTACACA AGCAGTAGTT  
 151 TCAGAAAGAA AAGTCATAGA AAAGCCTGTA GTCGAGAAAG TGCCTAGCCG  
 10 201 TCCTATCGCT AAAGAGACTC TAGCTGCACA GTTTATTGAA AGTAAGCCGG  
 251 TTATTGTAAAC CACACCACCC GTGCCTGTTG TTAGCGAAAC CCCAGAAAGTG  
 301 CCTACTGTGG CAGTTCGCC TCAGCCTGTT CGTGAGACAG TAAAAGAGGA  
 351 ACAAGCTCCT TATGCTACTG TTGTAGTGAA AAAAGGAGAT TTTCTCGAAC  
 401 GCATTGCGAG AGCAAATCAT ACTACCGTTG CAAAATTGAT GCAGATCAAT  
 15 451 GATCTTACCA CCACCCAAC TAAAAATGGT CAGGTCATCA AAGTCCCTAC  
 501 GTCTCAAGAT GTCAGCAACG AAAAACTCC TCAAACACAG ACCGCAAAAC  
 551 CTGAAATTA TTATATCGTC CAAGAAGGGG ATAGCCCGTG GACAATAGCA  
 601 TTGCGTAACC ATATTCGATT GGATGATTG CTAATAATGA ATGATCTCGA  
 20 701 TGAATATAAA GCCCGCGGCC TTAAGCCTGG AGATCAGTTG CGCATACGTT  
 GA

The PSORT algorithm predicts periplasmic (0.935).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 84A) and also in his-tagged form. The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (Figure 84B) and for FACS analysis.

25 These experiments show that cp7222 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 85

The following *C.pneumoniae* protein (PID 4377225) was expressed <SEQ ID 169; cp7225>:

30 1 MKGTPQYHFI GIGGIGMSAL AHILLDRGYE VSGSDLYESY TIESLKAKGA  
 51 RCFSGHDSH VPHDAVVVYS SSIAPDNVEY LTAIQRSSRL LHRÆLLSQL  
 101 MEGYESILVS GSHGTGTSS LIRAIQEAQ KDPSTYAGGL AANCLNGYSG  
 151 SSKIFVAREAD ESDGSLKHYT PRAVVITNID NEHLNNYAGN LDNLVQVIQD  
 201 PSRKVTDLNLK VFYNGDCPIL KGNVQGISYG YSPQCQLHIV SYNQKAWQSH  
 35 251 PSFTFLGQY QDIELNLPQG HNAANAAAAC GVALTFGIDI NIIRKALKKF  
 301 SGVHRLERK NISESFLFLE DYAHHPVEVA HTLRSVRDAV GLRRVIAIFQ  
 351 PHRFSRLEEC LQTFPKAFQE ADEVILTDVY SAGESPREST ILSDLAEQIR  
 401 KSSYVHCCYV PHGDIVDYLK NYIRIHDVCV SLGAGNIYTI GEALKDFNPK  
 451 KLSIGLVCGG KSCEHDISLL SAQHVSKYIS PRFYDVSYFI INRQGLWRTG  
 501 KDFPHLIEET QGDSPLSSEI ASALAKVDCL FVVLHGPFGE DGTIQGFFEI  
 40 551 LGKPYAGPSL SLAATAMDKL LTKRIASAVG VPVVPYQPLN LCFWKRNPKL  
 601 CIONLIETFS FPMIVKTAHL GSSIGIFLVR DKEELQEKIS EAFLYDIDVF  
 651 VERSRLGSRE IEVSCIGHSS SWYCMAGPNE RCGASGFIDY QEKYGFIDGID  
 701 CAKISFDLQL SQESLDCVRE LAERVYRAMQ GKGSARIDFF LDEEGNYWLS  
 751 EVNPIPGMTA ASPFLQAFVH AGWTQEQIVD HFIIDALHKF DRQQTIEQAF  
 45 801 TKEQDLVKR\*

The cp7225 nucleotide sequence <SEQ ID 170> is:

1 ATGAAGGGAA CTCCTCAGTA TCATTTTATC GGTATCGGTG GTATAGGAAT  
 51 GAGCGCTTTA GCTCATATTT TGCTTGATCG TGGCTATGAG GTCTCTGGAA  
 101 GCGACTTATA TGAAAGCTAT ACGATCGAAA GCCTGAAAGC TAAAGGTGCG  
 50 151 AGGTGTTTCT CAGGCCATGA TTCCTCCCAT GTTCTCATG ATGCCGTCGT  
 201 TGTATTATAGC TCAAGTATAG CCCCTGATAA TGTAGAGTAT CTTACCGCTA  
 251 TTCAAAGATC ATCAGTCTTT CTTCATAGAG CAGAGCTCTT GAGTCAGCTT  
 301 ATGAGAGGTT ATGAAAGCAT TCTGGTTTCA GGAAGCCATG GGAAGACAGG  
 351 GACCTCATCT CTAATTTCGAG CGATTTTCCA GGAAGCTCAG AAAGATCCCT

-124-

401 CCTATGCTAT TGGAGGACTC GCTGCAAACT GCCTGAATGG GTATTCTGGA  
 451 TCATCGAAAA TCTTCGTTCG CGAAGCCGAT GAAAGTGATG GGTCTTTAAA  
 501 GCACCTACACT CCCCGTGCAG TAGTCATTAC AAATATAGAT AATGAACATT  
 551 TGAATAATTA CGCTGGGAAT CTTGATAACC TGGTTCAGGT AATCCAGGAC  
 5 601 TTCTCTAGAA AAGTAACAGA TCTCAATAAG GTATTCTATA ACGGGGATTG  
 651 TCCTATTTTG AAAGGAAATG TCCAAGGGAT TTCTTATGGA TATTCACCAG  
 701 AATGTCAATT GCATATCGTT TCCTATAATC AAAAGGCATG GCAATCTCAC  
 751 TTTTCCTTTA CCTTTTATAG CCAGGAGTAT CAAGACATTG AGCTCAATCT  
 801 CCCTGGACAA CATAACGCTG CAAATGCAGC AGCAGCCTGT GGAGTTGCTC  
 10 851 TTACCTTTGG CATAGACATA AACATCATTG GAAAAGCTCT CAAAAAATTC  
 901 TCGGGAGTTC ATCGACGTCT AGAAAGAAAA AATATATCCG AAAGCTTTCT  
 951 TTTCTTAGAA GATTATGCTC ATCATCCTGT AGAGGTTGCA CATACCCTGC  
 1001 GCTCTGTGCG TGATGCTGTG GGTTCGCGAA GAGTCATCGC AATTTTTCOA  
 1051 CCACATCGAT TCTCTCGTTT AGAAGAGTGC TTACAAAACCT TCCCCAAAGC  
 15 1101 TTTCCAAGAA GCTGATGAAG TCATACTTAC AGATGTCTAT AGTGC CGGAG  
 1151 AAAGTCCTAG AGAGTCTATC ATTCTTTCCG ACCTTGCGGA ACAGATTTCGT  
 1201 AAGTCTTCTT ATGTCCATTG TTGTTATGTT CCCCATGGAG ACATCGTAGA  
 1251 TTATCTACGA AACTACATTG GCATTATGA TGCTGTGTT TCTCTAGGAG  
 1301 CTGGAAATAT CTATACTATT GGAGAGGCTT TAAAAGACTT TAACCCTAAA  
 20 1351 AAATATCCA TAGGACTCGT CTGTGGAGGG AAATCTGCG AACACGATAT  
 1401 TTCTCTACTT TCTGCTCAAC ATGTCTCTAA ATATATTCTT CCTGAATTCT  
 1451 ATGATGTGAG TTACTTCATC ATAAATCGTC AGGGCTTATG GAGAACAGGA  
 1501 AAGGATTTTC CTCATCTTAT TGAAGAGACT CAAGGGGATT CGCCACTTTC  
 1551 TTCTGAAATC GCTTCAGCTT TAGCAAAAGT CGACTGTTTG TTTCCCGTGC  
 25 1601 TCCATGGCCC ATTTGGAGAG GATGGTACGA TCCAGGGATT TTTTGAAATC  
 1651 TTAGGAAAAC CTTATGCCGG ACCCTCACTA TCTTTAGCAG CAACTGCAAT  
 1701 GGATAAGCTG TTAACAAAAC GAATGTCATC AGCAGTGGGT GTTCCTGTAG  
 1751 TCCCTTACCA ACCTTTAAAT CTCTGTTTCT GGAACGCAA TCCAGAACTA  
 1801 TGTATTCAGA ATCTTATAGA GACATTTTCT TTCCCTATGA TTGTAAAAAC  
 30 1851 TGCACATTG GATCTAGTA TTGGGATATT TTTAGTCCGT GATAAAGAGG  
 1901 AATTACAAGA AAAGATCTCA GAAGCATTTT TATATGACAC GGATGTGTTT  
 1951 GTGGAGGAAA GTCGCTTAGG GTCTCGTGAA ATCGAAGTGT CCTGTATCGG  
 2001 CCATTTCTCT AGCTGGTATT GTATGGCAGG GCCTAATGAA CGCTGTGGTG  
 2051 CTAGTGGGTT TATTGATTAT CAAGAGAAAT ATGGATTGTA TGGCATAGAT  
 35 2101 TGCGCAAAGA TCTCTTTTGA TTACAGCTC TCACAAGAAT CTTTAGATTG  
 2151 TGTTAGAGAA CTTGCAGAGC GTGTCTACCG AGCAATGCAA GGAAGGTT  
 2201 CAGCTCGAAT AGATTTTTC TTGGATGAAG AGGGGAATTA TTGGTTGTCA  
 2251 GAGGTCAATC CTATTCAGG AATGACAGCA GCTAGCCCAT TTTTACAAGC  
 2301 TTTTGTTCAC GCAGGATGGA CGCAAGAACA AATTGTAGAT CACTTTATTA  
 40 2351 TAGATGCTCT ACATAAGTTT GATAAGCAGC AGACTATCGA ACAGGCATTG  
 2401 ACTAAGAAC AAGATTTAGT TAAAAGATAA

The PSORT algorithm predicts inner membrane (0.16).

The protein was expressed in *E.coli* and purified as a his-tag product (Figure 85A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 85B) and for FACS analysis.

These experiments show that cp7225 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

#### Example 86

The following *C.pneumoniae* protein (PID 4377248) was expressed <SEQ ID 171; cp7248>:

50 1 MKFWLQGCAF VGCLLLTLPC CAARRRASGE NLQOTRPIAA ANLOWESYAE  
 51 ALEHSKQDHR PICLFFTGSD WCMWCIKMD QILQSSEFKH FAGVHLHMVE  
 101 VDFPQKNHQP EEQRQKNQEL KAQYKVTGFP ELVFDIAEGK QLARMGFEPG  
 151 GGAAYVSKVK SALKLR\*

A predicted signal peptide is highlighted.

55 The cp7248 nucleotide sequence <SEQ ID 172> is:

1 ATGAAATTTT GGTGCAAGG ATGTGCTTTT GTCGGTTGTC TGCTATTGAC

-125-

```

51   TTTACCTTGT TGTGCTGCAC GAAGACGTGC TTCTGGAGAA AATTTGCAAC
101  AAACTCGTCC TATAGCAGCT GCAAATCTAC AATGGGAGAG CTATGCAGAA
151  GCTCTTGAAC ATTCTAAACA AGATCACAAA CCTATTTGTC TTTTCTTTAC
201  AGGATCAGAC TGGTGTATGT GGTGCATAAA AATGCAAGAC CAGATTTTGC
5   251  AAAGCTCTGA GTTTAAGCAT TTTGCGGGTG TGCATCTGCA TATGGTTGAA
301  GTTGATTTCC CCCAAAAGAA TCATCAACCT GAAGAGCAGC GCCAAAAAAA
351  TCAAGAACTG AAAGCTCAAT ATAAAGTTAC AGGATTCCCC GAACTGGTCT
401  TCATAGATGC AGAAGGAAAA CAGCTTGCTC GCATGGGATT TGAGCCTGGT
10  451  GGTGGAGCTG CTTACGTAAG CAAGGTGAAG TCTGCTCTTA AACTACGTTA
501  A

```

The PSORT algorithm predicts periplasmic (0.932).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 86A) and also in his-tagged form. The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (Figure 86B) and for FACS analysis.

15 The cp7248 protein was also identified in the 2D-PAGE experiment.

These experiments show that cp7248 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

#### Example 87

The following *C.pneumoniae* protein (PID 4377249) was expressed <SEQ ID 173; cp7249>:

```

20   1  MIPSPFPINF RDDTILETDP KPSLIMFSSK KTEIASERRK AHPTLFKVLG
51   TIWNIVKFII SIILFLPLAL LWVLKKTQCF FILPSSIISQ SMSKTAVAIR
101  RMTFLSHIKQ LLSLKEISAA DRVVIQYDDL VVDSLAIKIP HALPHRWILY
151  SQGNSGLMEN LFDRGDSSLH QLAKATGSNL LVFNYPGIMS SKGEAKRENL
201  VKSYQACVRY LRDEETGPKA NQIIAFGYSL GTSVQAAALD REVTDGSDGT
25  251  SWIVVKDRGP RSLADVANI CKPIASAIK LVGWNIDSVK PSERLRCPFI
301  FIYNSNHDQE LISDGLFERE NCVATPFLEL PEVKTSGETKI PIPERDLLHL
351  NPLSPNVVDR LAAVISNYLD SENRKSQQPD *

```

The cp7249 nucleotide sequence <SEQ ID 174> is:

```

30   1  ATGATCCCAT CCCCTACCCC AATAAACTTT CGTGATGATA CGATTCTAGA
51   GACGGATCCA AAGCCGTCTT TAATCATGTT CTCTTCAAAA AAAACAGAGA
101  TAGCTTCTGA AAGACGGAAG GCCCATCCCA CCTTATTTAA AGTTCTAGGA
151  ACGATTTGGA ATATTTGTGA GTTTATTATC TCAATCATTC TGTTCCTTCC
201  CTTAGCGTTA TTGTGGGTAC TCAAGAAAAC CTGTCAGTTT TTCATTCTCC
25  251  CATCTTCTAT CATATCTCAG AGCATGTCAA AAACAGCTGT GGCAATTCGG
301  CGAATGACCT TTCTGTCCCA TATTAAACAA CTCCTAAGCC TTAAGGAAAT
35  351  CTCAGCTGCC GATCGTGTGG TTATACAATA TGACGATTIG GTGGTTGATA
401  GCTTAGCTAT AAAGATACCT CATGCTCTTC CCCACAGGTG GATTCTTTAT
451  TCTCAAGGAA ACTCTGATT GATGGAAAAC CTGTTTCGATC GGGGCGATTG
501  CTCCTACAC CAGCTAGCCA AAGCAACCGG CTCGAATCTT CTTGTGTTCA
55  551  ACTATCCTGG AATTATGTCC AGCAAAGGAG AAGCGAAACG AGAAATCTG
601  GTTAAATCGT ATCAGGCATG CGTACGCTAC CTACGAGATG AAGAGACAGG
651  TCCTAAAGCC AATCAAATCA TAGCTTTCGG ATACTCTTTG GGAACTAGTG
701  TCCAAGCTGC TGCTCTAGAT CGTGAGGTCA CTGATGGCAG TGATGGAAC
45  751  TCATGGATTG TTGTAAAAGA TCGGGGCCCT CGCTCTCTAG CAGATGTGCG
801  GAATCAAATT TGTAAGCCCA TAGCTTCCGC GATTATAAAA CTCGTGGTGT
851  GGAACATAGA CTCTGTGAAA CCTAGCGAAA GATTGCGTTG TCCCGAAATT
901  TTCATTTACA ACTCTAATCA TGATCAAGAA CTCATTAGCG ACGGCCTCTT
951  CGAAAGAGAA AATTGCGTAG CAACACCTTT TCTAGAGCTT CCTGAAGTAA
1001 A AACCTCGGG GACTAAAATT CCTATACCCG AAAGGATCT TCTCCATCTA
50  1051 AATCCTCTCA GTCCAAATGT AGTAGACAGA TTAGCAGCAG TGATCTCTAA
1101 TTATTTAGAT TCTGAAAACA GAAAGTCTCA GCAACCTGAT TAA

```

The PSORT algorithm predicts inner membrane (0.571).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 87A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 87B) and for FACS analysis.

These experiments show that cp7249 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

#### Example 88

The following *C.pneumoniae* protein (PID 4377261) was expressed <SEQ ID 175; cp7261>:

```

1  MLPISILLFY VILGCLSAYI ADKKKRNVIW WFFAGAFFGF IGLVVLILLP
51  SRRNALEKPO NDPFDNSDLF DDLKKSAGN DEIPSSGDLQ EIVIDTEKWF
101 YLNKDRENVG PISFEELVVL LKGTKYPEEI WVKRGMKDW QRVKDVPSLQ
151 QALKEASK*
```

The cp7261 nucleotide sequence <SEQ ID 176> is:

```

1  ATGCTCCCTA TTTGATTTT ATTATTTTAT GTGATTCTAG GTTGTCTATC
51  TGCCTACATA GCAGATAAGA AAAAACGAAA TGTATTTGGC TGGTTTITTTG
101 CAGGAGCATT TTTTGGATTT ATTGGTCTAG TTGTCCTTCT TCTTCTTCCT
151 TCTCGTCGAA ACGCTTTAGA AAAGCCACAA AACGATCCTT TTGATAACTC
201 CGATCTTTTT GATGATTTGA AAAAAAGTTT AGCAGGTAAT GACGAGATAC
251 CCTCATCGGG AGATCTTCAA GAAATCGTTA TCGATACAGA GAAGTGGTTT
301 TATTTAAATA AAGATAGAGA AAACGTAGGT CCGATATCTT TTGAGGAGTT
351 GGTCGTACTT TTAAGGGGAA AAACGTATCC AGAAGAAATT TGGGTATGGA
401 AAAAGGGAAT GAAAGATTGG CAACGAGTGA AGGATGTTC ATCACTACAA
451 CAGGCTTTGA AAGAAGCATC AAAATAA
```

The PSORT algorithm predicts inner membrane (0.848).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 88A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 88B) and for FACS analysis.

These experiments show that cp7261 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

#### Example 89

The following *C.pneumoniae* protein (PID 4377305) was expressed <SEQ ID 177; cp7305>:

```

1  MEVYSFHPAV RTSFQHRVMA ALDAWFFLGG HRLKVVSLDS CNSGWAYQEL
51  VSISTTEKVL KLLSYLLVPI VIIALLIRCL LHSNFRIDVE KERWLKIREL
101 GIDIESCKLP SSVVNQVSSF IWFEKDKSKR PRIDVDYHTL HSKDWVVPPI
151 VFQKIPKTSR FSYWFSQKET RKRDYVRNML DHVIGYLTSE GGEWLQYISK
201 TSYQSATSLD PERVLQYCLT DNQELQGEVQ RLLNERSATK SSGDKKVVLS
251 HVSDIICQCW WPKFLEVIQS PAFIEELVEE VSGKLNLDPL CLEKANTLDQ
301 ELRNSLLRAV VHHGSEGVDI KVGAGLIY TEAIQLQIPF SRS*
```

The cp7305 nucleotide sequence <SEQ ID 178> is:

```

1  ATGGAAGTTT ATAGTTTTC AACTGCGGTA AGGACTTCGT TTCAGCACCG
51  TGTAATGGCA GCACTAGATG CTTGGTTTTC TCTAGGAGGG CACCGTTTAA
101 AAGTAGTTTC TCTAGATAGT TGTAACCTCAG GTTGGGCGTA TCAAGAAGTT
151 GTGTCTATTT CAACGACAGA AAAAGTCTTG AAACCTACTCT CTTACCTACT
201 CGTACCGATT GTCATAATAG CTCTGTTAAT TCGTTGTCTT TTACATAGCA
251 ATTTTAGGAT AGACGTAGAG AAGGAACGTT GGTAAAAAAT AAGGGAGTTA
301 GGAATTGATA TAGAAAGCTG CAAACTCCCC AGTTCCTTATG TAAACCAGGT
351 TTCTCTGTTT ATTTGTTTG AAAAAGATAA ATCCAAACGG CCACGTATTG
401 ATGTAGATTA TCATACGCTA CATAGCAAAG ACTGGGTAGT TTTCCCTATC
```

-127-

451 GTTTTTCAGA AAATCCAAA GACCTCGCGT TTCAGTTATT GGTTCACACA  
 501 AAAAGAAACA AGGAAGAGGG ATTATGTGAG AAATATGCTG GACCACGTCA  
 551 TTGTTTATCT AACGTCAGAA GGTGGGGAGT GGTTCAGTA TATATCGAAA  
 5 ACCTCTTATC AAAGCGCTAC TTCTTTGGAT CCTGAAAGAG TTCTTCAATA  
 651 TTGCTTAACT GATAACCAGG AGCTCCAGGG AGAAGTGCAA CGTTTGCTTA  
 701 ATGAGGAGAG TCGGACCAA AGCTCTGGGG ATAAGGAAGT TTTGTTAAGT  
 751 CATGTATCTG ACATTATTTG CCACTGTTGG TGGCCAAAGT TTCTTGAAGT  
 801 TATACAATCT CCGGCCTTTA TTGAGAATT AGTAGAAGAA GTGAGTGGTA  
 851 AACTTAATTT AGATTTTTTA TGCCTAGAAA AGGCTAATAC ATTAGATCAG  
 10 GAGTTGAGAA ACAGTCTTCT AAGAGCAGTC GTACACCACG GTTCTGAAGG  
 951 AGTTGATATT AAGAAAGTTG GTGCCGGCCT CATTATTTAT ACGGAAGCTA  
 1001 TTCAATTACA GATTCCCTTC TCAAGGAGTT AA

The PSORT algorithm predicts inner membrane (0.508).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 89A) and also as a  
 15 double GST/his fusion. The recombinant proteins were used to immunise mice, whose sera were  
 used in a Western blot (Figure 89B) and for FACS analysis.

These experiments show that cp7305 is a surface-exposed and immunoaccessible protein, and that it  
 is a useful immunogen. These properties are not evident from the sequence alone.

#### Example 90

20 The following *C.pneumoniae* protein (PID 4377347) was expressed <SEQ ID 179; cp7347>:

1 MKKGLGAIIV FGLLFTSSVA GFSKDLTKDN AYQDLNVIEH LISLKYAPLP  
 51 WKELLFGWDL SQQTQQARLQ LVLEEKPTTN YCQKVLNRYV RSLNDYHAGI  
 101 TFYRTESAYI PYVLKLSERG HVFVVDVQTS QGDIYLGDEI LEVDGMGIRE  
 151 AIESLRFGRG SATDYSAAVR SLTSRSAAFG DAVPSGIAML KLRRPSGLIR  
 201 STFVRWRYTP RHIGDPSLVA PLIPEHKPQL PTQSCVLFPS GVNSQSSSSS  
 251 LFSSYMPYYP WEELRVQNKQ RFDNHHIGS RNFGLPTFGP ILWEQDKGPY  
 301 RSYIFKAKDS QGNPHRIGFL RISSYVWTDL EGLEBDHKDS PWELFGEIID  
 351 HLEKETDALI IDQTHNPGRS VFYLYSLLSM LTDHPLDTFK HRMIFQDEV  
 401 SSALHWQDLL EDVFTDEQAV AVLGETMEGY CMDMHAVASL QNFSQSVLSS  
 30 WSGDINLSK FPLLGFPAQV RPHPKQYTK PLFMLIDEDD FSCGDLAPAI  
 501 LKDNCRATLI GKPTAGAGGF VFQVTFPNRS GIKGLSLTGS LAVRKDGEFI  
 551 ENLGVAPHID LGFTSRDLQT SRFTDYVEAV KTIVLTSLSE NAKKSEQTS  
 601 PQETPEVIRV SYPTTTSAS\*

A predicted signal peptide is highlighted.

35 The cp7347 nucleotide sequence <SEQ ID 180> is:

1 ATGAAAAAAG GGAAATTAGG AGCCATAGTT TTTGGCCTTC TATTTACAAG  
 51 TAGTGTGCT GGTTTTCTA AGGATTGAC TAAAGACAAC GCTTATCAAG  
 101 ATTTAAATGT CATAGAGCAT TTAATATCGT TAAAAATATG TCCTTTACCA  
 151 TGGAAAGAAC TATTATTTGG TTGGGATTTA TCTCAGCAA CACAGCAAGC  
 40 TCGCTGCAA CTGGTCTTAG AAGAAAAACC AACAACCAAC TACTGCCAGA  
 251 AGGTACTCTC TAACTACGTG AGATCATTA ACGATTATCA TGCAGGGATT  
 301 ACGTTTATC GTACTGAAAG TCGGTATATC CCTTACGTAT TGAAGTTAAG  
 351 TGAAGATGGT CATGCTTTG TAGTCGACGT ACAGACTAGC CAAGGGGATA  
 401 TTTTACTTAGG GGATGAAATC CTTGAAGTAG ATGGAATGGG GATTCTGTAG  
 45 GCTATCGAAA GCCTTCGCTT TGGACGAGGG AGTGCCACAG ACTATTCTGC  
 501 TGCAGTTCGT TCCTTGACAT CGCGTTCCGC CGCTTTTGGA GATGCGGTTC  
 551 CTTCAGGAAT TGCCATGTTG AAACCTCGCC GACCCAGTGG TTTGATCCGT  
 601 TCGACACCGG TCCGTTGGCG TTATACTCCA GAGCATATCG GAGATTTTTC  
 651 TTTAGTTGCT CCTTTGATTC CTGAACATAA ACCTCAATTA CCTACACAAA  
 50 GTTGTGTGCT ATTCCGTTCC GGGGTAAATT CACAGTCTTC TAGTAGCTCT  
 751 TTATTAGATT CCTACATGGT GCCTTATTTT TGGGAAGAA TCGGGGTICA  
 801 AAATAAGCAG CGTTTGTACA GTAATCACCA TATAGGGAGC CGTAATGGAT  
 851 TTTTACCTAC GTTTGGTCCT ATTCTTTGGG AACAAGACAA GGGGCCCTAT  
 901 CGTTCCCTATA TCTTTAAAGC AAAAGATTCT CAGGGCAATC CCCATCGCAT  
 55 AGGATTTTTA AGAATTTCTT CTTATGTTTG GACTGATTTA GAAGGACTTG  
 1001 AAGAGGATCA TAAGGATAGT CCTTGGGAGC TCTTTGGAGA GATCATCGAT

-128-

1051 CATTTGGAAG AAGAGACTGA TGCTTTGATT ATTGATCAGA CCCATAATCC  
 1101 TGGAGGCAGT GTTTTCTATC TCTATTCGTT ACTATCTATG TTAACAGATC  
 1151 ATCCTTTAGA TACTCCTAAA CATAGAATGA TTTTCACTCA GGATGAAGTC  
 1201 AGCTCGGCTT TGCAGTGGCA AGATCTACTA GAAGATGTCT TCACAGATGA  
 1251 GCAGGCAGTT GCCGTGCTAG GGGAAACTAT GGAAGGATAT TGCATGGATA  
 1301 TGCATGCTGT AGCCTCTCTT CAAAACCTCT CTCAGAGTGT CCTTTCTTCC  
 1351 TGGGTTTCAG GTGATATTAA CCTTTCAAAA CCTATGCCCT TGCTAGGATT  
 1401 TGCACAGGTT CGACCTCATC CTAAACATCA ATATACTAAA CCTTTGTTTA  
 1451 TGTGTATAGA CGAGGATGAC TTCTCTGTGT GAGATTTAGC GCCTGCAATT  
 1501 TTGAAGGATA ATGGCCGCGC TACTCTCAT TGGAAAGCCAA CAGCAGGAGC  
 1551 TGGAGGTTT GTATTCCAAG TCACCTTCCC TAACCGTTCT GGAATTAAAG  
 1601 GTCTTTCCTT AACAGGATCT TTAGCTGTGA GGAAGATGG TGAGTTTATT  
 1651 GAAACTTAG GAGTGGCTCC TCATATTGAT TTAGGATTTA CCTCCAGGGA  
 1701 TTTTGCAAACT TCCAGGTTTA CTGATTACGT TGAGGCAGTG AAAACTATAG  
 1751 TTTTAACTTC TTTGTCTGAG AACGCTAAGA AGAGTGAAGA GCAGACTTCT  
 1801 CCGCAAGAGA CGCCTGAAGT TATTCGAGTC TCTTATCCCA CAACGACTTC  
 1851 TGCTTCGTAA

The PSORT algorithm predicts periplasmic space (0.2497).

The protein was expressed in *E. coli* and purified as a GST-fusion product (Figure 90A) and also in  
 his-tagged form. The recombinant proteins were used to immunise mice, whose sera were used in a  
 Western blot (Figure 90B) and for FACS analysis.

These experiments show that cp7347 is a surface-exposed and immunoaccessible protein, and that it  
 is a useful immunogen. These properties are not evident from the sequence alone.

#### Example 91

The following *C. pneumoniae* protein (PID 4377353) was expressed <SEQ ID 181; cp7353>:

1 MNMPVPSAVP SANITLKEDS STVSTASGIL RTATGEVLVS CTALEGSST  
 51 DALISLALGQ IILATQOELL LQSTNVHQLL FLPPEVVELE IQVVDLLVQL  
 101 EHAETITSEP QETQTSRSE QTLPOQSSSK QSALSPRSLK PEISDSKQQQ  
 151 ALQTPKDSAV RKHSEAPSPE TQARASLSQA SSSQSRSLPP QESAPERTLL  
 201 EQQKASSFSP LSQPSAEKQK EALTTSKSHE LYKERDQDRQ QREQHDRKHD  
 251 QEEDAESKKK KKKRGLGVEA VAEPPGENLD IALIFSDQM RPPAETS SKK  
 301 ETTFFKKLPS PMSVPSRFIP SKNPLSVGSS IHGPIQTPKV ENVFLRFMKL  
 351 MARILQARA EANELYMRVK QRTDDVDTLT VLISKINNER KDIDWSENER  
 401 MKALLNRAKE IGVTDKEKY TWTREEKRL KENVQMRKEN MEKITQMERT  
 451 DMQRHLQEIS QCHQARSNVL KLLKELMDTF IYNLRP\*

The cp7353 nucleotide sequence <SEQ ID 182> is:

1 ATGAATATGC CTGTTCCCTC TGCAGTCCC TCTGCAAATA TAACTCTAAA  
 51 AGAAGACAGC TCAACAGTTT CCACAGCCTC TGGAAATATTA AAGACTGCAA  
 101 CAGGTGAAGT CTTAGTCTCT TGTACAGCGC TAGAAGGAAG CTCTTCTACA  
 151 GATGCTTTAA TTAGCTTAGC TTTAGGACAA ATCATTCTTG CGACCCAACA  
 201 AGAAGTGTCT TTACAAAGCA CAAATGTTCA TCAACTCCTC TTCTCCCTC  
 251 CTGAAGTTGT AGAATTAGAA ATCCAAGTTG TTGACTTGCT AGTGCAATTG  
 301 GAACATGCGAG AGACAATCAC AAGTGAACCA CAAGAAACAC AAACGCAAAG  
 351 TAGGAGTGAG CAGACCCTCC CTCAACAAAG CAGCAGTAAA CAATCTGCTC  
 401 TCTCCCCACG CTCCTTAAAA CCTGAAATTT CTGATTCTAA ACAACAGCAA  
 451 GCTCTTCAAA CACCAAAAGA CTCTGCTGTA AGAAAACACA GCGAAGCACC  
 501 GTCACCTGAG ACACAAGCTC GCGCTTCCTT ATCTCAGGCA AGCTCAAGTT  
 551 CTCAGAGATC CTTACCTCCG CAAGAAAGTG CGCCAGAAAG AACACTATTA  
 601 GAACAACAAA AAGCAAGCTC CTTCTCTCCT CTATCCAGT TCTCTGCAGA  
 651 GAAACAAAAA GAGGCCCTGA CGACCTCAAA ATCTCATGAA CTCTATAAAG  
 701 AACGCGATCA AGATCGCCAA CAAAGAGAGC AGCAGCAGAG AAAGCAGCAT  
 751 CAGGAAGAAG ACGCTGAATC TAAAAAGAAA AAGAAGAAAC GTGGTCTCGG  
 801 TGTAGAGGCA GTCGCTGAGG AACC CGGAGA AAATCTAGAT ATTGCCGCTT  
 851 TAATCTTCTC AGATCAAATG CGACCTCCTG CTGAAGAAAC TTCTAAAAAA  
 901 GAAACGACAT TCAAAAAGAA GCTACCTTCT CCAATGTCTG TGTTTAGCAG  
 951 ATTCAATCCCT AGTAAGAATC CGTTATCTGT AGGCTCTTCA ATACACGGGC  
 1001 CTATACAAAC TCCAAAAGTA GAAAATGTGT TCTTAAGGTT CATGAAGCTC

-129-

```

1051 ATGGCAAGAA TCTTAGGCCA AGCCGAAGCC GAAGCTAATG AACTCTACAT
1101 GCGAGTCAAA CAACGTACCG ATGATGTAGA CACACTCACA GTCCTTATCT
1151 CTAAGATCAA TAATGAAAAG AAAGACATTG ATTGGAGTGA AAATGAAGAG
1201 ATGAAAGCTC TTTTAAATCG AGCTAAAGAG ATTGGAGTCA CTATAGACAA
1251 AGAAAAATAT ACTTGGACAG AAGAGGAAAA AAGACTTCTA AAAGAGAATG
1301 TCCAAATGCG CAAAGAGAAAT ATGGAGAAAA TCACTCAAAT GGAAAGGACG
1351 GACATGCAAA GGCACCTCCA AGAGATTCTT CAATGTCATC AAGCGCGCTC
1401 TAATGTATTG AAGTTATTGA AAGAACTTAT GGACACCTTC ATTTACAACC
1451 TACGCCCCTA A

```

10 The PSORT algorithm predicts cytoplasm (0.1308).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 91A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 91B) and for FACS analysis.

15 These experiments show that cp7353 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 92

The following *C.pneumoniae* protein (PID 4377408) was expressed <SEQ ID 183; cp7408>:

```

20 1 MLKIQKKRMC VSVVITVGAI VGFNSADAA PKKKKIPIQI LYSFTKVSSY
51 LKNEASTIF CVDVDRGLLQ HRYLGSPGWQ ETRRRQLFKS LENQSYGNER
101 LGERTLAIDI FRNKECLESE IPEQMEAILA NSSALVLGIS SFGITGIPAT
151 LHSLLRQNL FQKRSIASES FLLKIDSAPS DASVFKGVL FRGETAIVDA
201 LSQLEAQLDL SPKKIIFLGE DPEVVQAVGS ACIGWGMNPL GLVYYPQES
251 LFSYVHPYST ATELQEAQGL QVISDEVAQL TLNALPKMN*

```

The cp7408 nucleotide sequence <SEQ ID 184> is:

```

25 1 ATGTTGAAAA TCCAGAAAAA AAGAATGTGT GTCAGCGTAG TCATCACGGT
51 AGGCCCCATA GTGGGGTTTT TCAATTCTGC AGACGCGACA CCAAAGAAAA
101 AGAAGATCCC TATACAGATT CTCTACTCCT TTAATAAAGT CTCTTCCTAT
151 TTAAAAACG AAGACGCAAG TACTATATTT TGCGTCGATG TGGATCGTGG
201 ACTTCTCCAG CATCGGTATT TAGGTAGTCC AGGATGGCAG GAAACCAGAC
30 251 GTCGCGAGTT ATTTAAATCC TTAGAAAATC AATCATACGG CAACGAACGT
301 TTAGGAGAAG AAACCTCTGC TATTGATATT TTCAGGAACA AAGAGTGCTT
351 GGAGAGCGAG ATCCAGAGC AGATGGAAGC TATCCTTGCA AATTCTCGG
401 CCTTGGTCTT AGGCATCTCT TCTTTTGGGA TCACAGGAAT TCCTGCGACT
451 TTGCATAGTT TGCTTCGACA GAATCTATCT TTCCAAAAAC GCTCTATAGC
35 501 ATCGGAGAGC TTCCTTTTAA AGATCGATAG TGCCCCCTCA GATGCCTCTG
551 TTTTATTATA AGCGGTGCTT TTCCGCGGAG AGACTGCGAT CGTGGATGCG
601 TTAAGCCAAT TATTTGCCCA GCTCGATCTT TCTCTAAAA AAATTATCTT
651 TCTAGGAGAA GACCCTGAGG TCGTTCAAGC TGTTGGGTCT GCTTGATATG
701 GTTGGGGCAT GAACTTTTTA GGCTGGTAT ACTATCCTGC TCAAGAAAGC
40 751 CTTTTTCTT ATGTTTCATCC TTAATCTACA GCAACGAGC TCCAAGAAGC
801 ACAGGTTTA CAAGTAATTT CAGATGAAGT CGCACAGCTT ACTTTAAACG
851 CTCTCCGAA AATGAATTAA

```

The PSORT algorithm predicts inner membrane (0.123).

45 The protein was expressed in *E.coli* and purified as a his-tag product (Figure 92A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 92B) and for FACS analysis.

These experiments show that cp7408 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.



**Example 93**

The following *C.pneumoniae* protein (PID 4376424) was expressed <SEQ ID 185; cp6424>:

```

1  MMHNIVVLSE  EPGRSAFLGR  TAFFPNKYPI  AQGGVGIPST  IGNLFTIWYC
5  51  FYFYRAATPQ  SDHPDGCIFI  LLERLKLGA  GFFYCDLRES  NTTGFTLFFE
101 GSNKGVLNH  LFIRDE*

```

The cp6424 nucleotide sequence <SEQ ID 186> is:

```

1  ATGATGCACA  ATATTGTTGT  TCTTAGTGAG  GAACCTGGAC  GAAGCGCTTT
10  51  TCTTGGTAGG  ACGGCATTTT  TCCCTAATAA  GTATCCAATA  GCTCAGGGTG
101 GTGTGGAAT  ACCATCTACA  ATAGGCAATC  TCTTTACTAT  ATGGTACTGT
151 TTCTATTTT  ATAGAGCTGC  AACTCCACAA  TCTGATCATC  CTGACGGATG
201 TGGCTTTAT  CTACTAGAAA  GGCTTAAGGA  GCTCGGTGCA  GGGTTCTTTT
251 ATTGATGAT  TCGTGAGTCC  AATACCACTG  GCTTTACTCT  TTTTGTGAA
301 GGCTCCAATA  AAGGTGTGTT  AAAGAATCAC  TTGTTTATTA  GAGATGAGTA
351 A

```

15 The PSORT algorithm predicts cytoplasm (0.2502).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 93A) and also in his-tagged form. The recombinant proteins were used to immunise mice, whose sera were used in Western blots (Figure 93B) and for FACS analyses (Figure 93C; GST-fusion).

20 These experiments show that cp6424 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

**Example 94**

The following *C.pneumoniae* protein (PID 4376449) was expressed <SEQ ID 187; cp6449>:

```

1  VASETYPYQI  LHAQREVRDA  YFNQADCHPA  RANQILEARK  ICLLDVYHTN
25  51  HYSVPTFCVD  NYPNLRFTFV  SSKNEMENGL  SNPLDNVLVE  AMVRRTHARN
101 LLAACKIRNI  EVPRVVGDL  RSGILISKLE  LKQPQFQSLT  EDFVNHSTNQ
151 EEARVHQKHV  LLISLILLCK  QAVLESFQEK  KRSS*

```

The cp6449 nucleotide sequence <SEQ ID 188> is:

```

1  GTGGCGTCTG  AAACGTATCC  TTCTCAGATA  TTGCACGCTC  AGAGGGAAGT
30  51  ACGTGATGCC  TATTTTAATC  AAGCGGATTG  CCATCCTGCT  CGGGCTAATC
101 AGATTCTCGA  GGCTAAGAAA  ATCTGTTTAT  TAGATGTTTA  TCATACTAAT
151 CATTATTCG  TATTACTTTT  TTGTGTAGAT  AATTATCCGA  ATCTCCGCTT
201 TACATTTGTA  TCTTCAAAAA  ACAATGAGAT  GAATGGCTTA  TCTAATCCTC
251 TAGATAATGT  TCTTGTAGAG  GCTATGGTAC  GTAGAACACA  TGCAAGAAAC
35  301  CTACTTGCAG  CGTGTAATAA  TCGAAATATT  GAGGTTCCAA  GGGTTGTTGG
351 GCTTGACCTA  AGATCTGGGA  TACTCATTTT  GAAACTAGAA  TTGAAGCAAC
401 CTCAGTTCCA  AAGTTTAACA  GAAGACTTCG  TAAATCATTC  CACAAATCAG
451 GAAGAAGCTC  GCGTCCATCA  AAAGCATGTG  TTGCTAATTT  CTTTAATTTT
501 ACTTTGCAAG  CAGGCCGTTC  TGAATCATT  CCAGGAAAAA  AAGCGATCCT
551 CTTAA

```

40 The PSORT algorithm predicts inner membrane (0.2084).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 94A) and also in his-tagged form. The recombinant proteins were used to immunise mice, whose sera were used in Western blots (Figure 94B) and for FACS analyses (Figure 94C; GST-fusion).

45 These experiments show that cp6449 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

**Example 95**

The following *C.pneumoniae* protein (PID 4376495) was expressed <SEQ ID 189; cp6495>:

MRELNAFELTQPEEYRNRRWVLMPCCLKRCRFTQHAKVWSYRCVHEASLYEKNCFLLTYDDKHLPOYGSVLVHLHLQLFLKR  
LRKMISPHKIRYFECGAYGFKLQRPYHLLLS

5 The cp6495 nucleotide sequence <SEQ ID 190> is:

TTGCGAGAATTAAATGCTTTTGAATTAACCTCAACCTGAAGAGTATCGAAACCGTTGGGTTTGTATGCCTTGCTTAAAGTGT  
CGTTTTGTAGAACGCAACATGCAAAAGTCTGGTCTTATCGTTGTGTCCATGAAGCTTCTTTGTATGAGAAAAATTGTTTT  
CTTACTTTGACTTATGATGATAAGCATTTACCTCAGTATGGTTCGTTGGTAAAGCTGCATTTACAGCTGTTCTTAAAGAGA  
TTAAGAAAGATGATTTCTCTCATAAATTCGTTATTTGAATGTGGTTCGTTATGGAACCAAATTACAAAGACCTCATTAT  
CATCTACTTTTATCATGA

10 The PSORT algorithm predicts cytoplasmic (0.280).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 95A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 95B) and for FACS analysis (Figure 95C).

15 These experiments show that cp6495 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

**Example 96**

The following *C.pneumoniae* protein (PID 4376506) was expressed <SEQ ID 191; cp6506>:

1 MRRFLFLILS SLPLVAFSAD NPTILEEKQS PLSRVSIIFA LPGVTPVSPD  
51 GNCPTLPWFH SKKTLLEGQRI YYSGDSFGKY FVVSALWPNK VSSAVVACNM  
101 ILKHRVDLIL IIGSCYSRSQ DSRFGSVLVS KGYINYDADV RPPFFERFPI  
151 DIKKSVPFATS EVHREAILRG GEEFISTHKQ EIEELLKTHG YLKSTTKTEH  
201 TLMBGLVATG ESPAMSRNYP LSLQKLYPEI HGFDSVSGAV SQVCYBYSIP  
251 CLGVNILLPH PLESRSNEDW KHLQSEASKI YMDTLKSVL KELCSSH\*

25 The cp6506 nucleotide sequence <SEQ ID 192> is:

1 ATGCGTCGTT TTCTGTTTCT TATTCTTAGC TCTCTTCCTT TGGTCGCATT  
51 CTCTGCTGAT AATTTCACTA TTCTAGAAGA AAAACAGAGT CCTTTAAGTC  
101 GTGTAAGTAT TATTTTGGCT TTACCTGGGG TTAACCTCCGT TTCTTTTGAT  
151 GGTAATTGTC CTATTCCTTG GTTTTCTCAT AGTAAAAAGA CTCTAGAGGG  
30 201 ACAGAGAATT TATTACTCTG GCGACTCCTT TGGGAAATAC TTTGTAGTTT  
251 CTGCTCTTTG GCCTAATAAA GTTCTTCAG CTGTTGTGGC TTGTAATATG  
301 ATTCTTAAAC ATCGAGTGGG TCTTATTCTA ATTATAGGCT CGTGTACTC  
351 TAGGTCTCAA GATAGCCGTT TTGGCAGCGT CTTAGTTTCT AAAGGCTACA  
401 TTAATTATGA TGCAGATGTG AGGCCTTTCT TTGAAGATT TGAGATTCCA  
35 451 GACATTAAAA AGAGTGTTTT TGCAACCACT GAGGTTTCATC GGGAGGCAAT  
501 TCTTCGTGGA GGCGAAGAGT TTATTTCTAC CCATAAACAA GAAATCGAAG  
551 AGCTTTTGAA GACTCATGGG TATTTGAAAT CAACAACCAA AACGGAGCAC  
601 ACCTTAATGG AAGGTTTGGT TGCTACAGGC GAGTCTTTCCG CGATGTCGCG  
651 AAATATTTTT CTTTCCCTTAC AAAAATTGTA TCCAGAGATT CATGGTTTTG  
40 701 ATAGTGTCAG CGGCGTGT TCTCAGGTAT GCTATGAATA TAGCATTCCT  
751 TGTTTAGGTG TGAATATCCT TCTCCCTCAT CCTTTAGAAAT CACGGAGTAA  
801 CGAGGATTGG AAGCATCTTC AAAGTGAGGC AAGTAAAT TATATGGATA  
851 CCTTGCTCAA GAGTGATTA AAAGAACTCT GTTCTTCTCA TTAA

The PSORT algorithm predicts periplasmic space (0.571).

45 The protein was expressed in *E.coli* and purified as his-tag (Figure 96A) and GST-fusion (Figure 96B) products. The GST-fusion protein was used to immunise mice, whose sera were used in a Western blot (Figure 96C) and for FACS analysis (Figure 96D).

These experiments show that cp6506 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 97

The following *C.pneumoniae* protein (PID 4376882) was expressed <SEQ ID 193; cp6882>:

```

5      1  MSLNLNPSSQ DSASEDSTSQ SQIFDPIRNR ELVSTPEEKV RQRLLSFLMH
      51  KLNYPKKLII IEKELKTLFP LLMRKGTLP KRRPDILII PPTYTDAQGN
     101  THNLGDPKPL LLIECKALAV NQNALKQLLS YNYSIGATCI AMAGKHSQVS
     151  ALFNPKTQTL DFYPGLPEYS QLLNYFISLN L*

```

The cp6882 nucleotide sequence <SEQ ID 194> is:

```

10      1  ATGTCCTTAT TGAACCTTCC CTCAAGCCAG GATTCTGCAT CTGAGGACTC
      51  CACATCGCAA TCTCAAATCT TCGATCCCAT TAGAAATCGG GAGTTAGTTT
     101  CTACTCCCGA AGAAAAAGTC CGCCAAAGGT TGCTCTCCTT CCTAATGCAT
     151  AAGCTGAAC T ACCCTAAGAA ACTCATCATC ATAGAAAAAG AACTCAAAAC
     201  TCTTTTTCCT CTGCTTATGC GTAAAGGAAC CCTAATCCCA AAACGCCGCC
15      251  CAGATATTCT CATCATCACT CCCCCACAT ACACAGACGC ACAGGGAAAC
      301  ACTCACAACC TAGGGGACCC AAAACCCCTG CTACTTATCG AATGTAAGGC
     351  CTTAGCCGTA AACCAAAATG CACTCAAACA ACTCCTTAGC TATAACTACT
     401  CTATCGGAGC CACCTGCATT GCTATGGCAG GGAACACTC TCAAGTGTCA
     451  GCTCTCTTCA ATCCAAAAC ACAAACTCTT GATTTTATC CTGGCCTCCC
20      501  AGAGTATTCC CAACTCCTAA ACTACTTTAT TTCTTTAAAC TTATAG

```

The PSORT algorithm predicts cytoplasm (0.362).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 97A). The protein was used to immunise mice, whose sera were used in a Western blot (Figure 97B) and for FACS analysis (Figure 97C).

25 These experiments show that cp6882 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 98

The following *C.pneumoniae* protein (PID 4376979) was expressed <SEQ ID 195; cp6979>:

```

30      1  MSVNPSPGNSK NDLWITGAHD QHPDVKESGV TSANLGSHRV TASGGRQGLL
      51  ARIKEAVTGF FSRMSFFRSQ APRGSQQPSA PSADTVRSPL PGGDARATEG
     101  AGRNLKKG Y QPGMKVTIPQ VPGGGAQRSS GSTTLKPTRP APPPPKTGGT
     151  NAKRPATHGK GPAPQPPEKTG GTNAKRAATH GKGPAPQPPK GILKQPGQSG
     201  TSGKKRVSW DED*

```

The cp6979 nucleotide sequence <SEQ ID 196> is:

```

35      1  ATGTCGTGTA ATCCATCAGG AAATTTCCAAG AACGATCTCT GGATTACGGG
      51  AGCTCATGAT CAGCATCCCG ATGTTAAAGA ATCCGGGGTT ACAAGTGCTA
     101  ACCTAGGAAG TCATAGAGTG ACTGCCCTCAG GAGGACGCCA AGGGTTATTA
     151  GCACGAATCA AAGAAGCAGT AACC GG GTTT TTTAGTCGGA TGAGCTTCTT
     201  CAGATCGGGA GCTCCAAGAG GTAGCCAACA ACCCTCTGCT CCATCTGCAG
40      251  ATACTGTACG TAGCCCGTTG CCGGGAGGGG ATGCTCGCGC TACCGAGGGA
      301  GCTGTGAGGA ACTTAATTA AAAAGGGTAC CAACCAGGGA TGAAAGTCAC
     351  TATCCACAG GTTCCTGGAG GAGGGGCCCA ACGTTCATCA GGTAGCACGA
     401  CACTAAAGCC TACGCGTCCG GCACCCCCAC CTCCTAAAAC GGGTGGAAC
     451  AATGCAAAAC GTCCGGCAAC GCACGGGAAG GGTCCAGCAC CCCAGCCTCC
45      501  TAAACAGGT GGGACCAATG CTAAGCGCGC AGCAACGCAT GGGAAAGGTC
      551  CAGCACCTCA ACCTCCTAAG GGCATTTTGA AACAGCCTGG GCAGTCTGGG
     601  ACTTCAGGAA AGAAGCGTGT CAGCTGGTCT GACGAAGATT AA

```

The PSORT algorithm predicts cytoplasm (0.360).

-133-

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 98A). The GST-fusion protein was used to immunise mice, whose sera were used in a Western blot (Figure 98B) and for FACS analysis (Figure 98C).

These experiments show that cp6979 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

#### Example 99

The following *C.pneumoniae* protein (PID 4377028) was expressed <SEQ ID 197; cp7028>:

```

1  MLLGFLCDCP CASWQCAAVA NCYDSVFMSR PEHKPNIPYI TKATRRGLRM
51  KTLAYLASLK DARQLAYDFL KDPGSLARLA KALIAPKEAL QEGNLPFYGC
101 SNIEDILEEM RRPHRILLIG FSYCQPKAC PEGRFNDACR YDPSHPTCAS
151 CSIGTMMRLN ARRYTTVIIIP TFIIDIAKHLH TLKKRYPGYQ ILFAVTACEL
201 SLKMFPGDYAS VMNLKGVGIR LTGRICNTFK AFKLAERGVK PGVTILEEDG
251 FEVLARILTE YSSAPFPRDF CEIH*
```

The cp7028 nucleotide sequence <SEQ ID 198> is:

```

1  ATGCTTCTAG GGTTTTGTG TGA CTGCCCC TGTGCTTCGT GGCAGTGTGC
51  GGCCGTTGCT AATTGTTATG ATTCCGTATT TATGCTCTAGA CCAGAGCACA
101 AACCTAATAT TCCTTATATT ACTAAAGCTA CAAGACGGGG TCTGCGTATG
151 AAGACGCTTG CTTATCTGGC CTCCTTAAAA GATGCTAGAC AGCTTGCCTA
201 TGATTCTCTG AAAGATCCTG GTTCTTTAGC TCGGTTAGCT AAGGCTTTGA
251 TAGTCCTTAA GGAGGCCTTA CAGGAGGGCA ACCTATTTTT TTATGGCTGT
301 AGTAATATTG AGGATATTTT AGAGGAGATG CGTCGTCCTC ATAGAATCCT
351 TTTGTTAGGA TTTTCTTATT GTCAAAAGCC TAAGGCATGT CCTGAAGGGC
401 GTTTCATAGA TGCTTGTCGG TATGATCCTT CACATCCTAC ATGTGCCTCA
451 TGTTCATATG GGACCATGAT GCGGCTGAAT GCTCGTAGAT ACACTACTGT
501 GATCATCCCT ACATTTATAG ATATCGCAAA ACATTTACAC ACTTTAAAAA
551 AGCGCTACCC TGGATATCAA ATTCTCTTTG CAGTTACTGC TTGTGAACTT
601 TCCTTAAAAA TGTTTGGAGA TTATGCCTCC GTAATGAACT TAAAGGGTGT
651 GGGCATCAGA CTCACAGGAC GTATTTCGCA TACATTTAAG GCATTTAAAT
701 TAGCTGAGCG AGGAGTCAAA CCAGGAGTCA CTATCCTAGA AGAAGATGGC
751 TTTGAGGTAT TAGCAAGGAT TCTTACAGAA TACAGTAGCG CTCCTTTCCC
801 TAGAGACTTT TGTGAGATCC ATTAG
```

The PSORT algorithm predicts cytoplasm (0.1453).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 99A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 99B) and for FACS analysis (Figure 99C).

These experiments show that cp7028 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

#### Example 100

The following *C.pneumoniae* protein (PID 4377355) was expressed <SEQ ID 199; cp7355>:

```

1  MKKVVTLSII FFATYCASEL SAVTVVAVPL SEAPGKIQVR PVVGLQFQEE
51  QGSVPYSFYY PYDYGYYYPE TYGYTKNTGQ ESRECYTRFE DGTIFYECD*
```

The cp7355 nucleotide sequence <SEQ ID 200> is:

```

1  ATGAAGAAAG TCGTAACACT ATCCATTATA TTTTTCGCAA CGTATTGTGC
51  ATCAGAGCTT AGTGCTGTAA CTGTAGTGGC TGTGCCTTTA TCAGAGGCTC
101 CAGGGAAGAT TCAAGTTCGT CCCGTCGTTG GTCTGCAATT TCAAGAAGAA
151 CAGGGTTCCTG TGCCCTATAG TTTTATTAT CCTTATGACT ATGGGTATTA
201 CTATCCAGAG ACTTATGGCT ATACTAAAAA TACAGGTCAA GAAAGTCGCG
```

-134-

251 AATGTTATAC CCGATTGAA GATGGCACAA TTTTATGA ATGCGATTAG

The PSORT algorithm predicts inner membrane (0.143).

The protein was expressed in *E.coli* and purified as a GST-fusion (Figure 100A) and a his-tag product. The proteins were used to immunise mice, whose sera were used in a Western blot (Figure 100B) and for FACS analysis (Figure 100C).

These experiments show that cp7355 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 101

The following *C.pneumoniae* protein (PID 4377380) was expressed <SEQ ID 201; cp7380>:

```

10      1  VHVCERTLDP KYILKIALKL RQSLSLFFQN SQSLQRAYST PYSYVRIILQ
      51  KENKEKQALA RHKCSISILEF FKNLLFVHLL SLSKNQREGC STDMAVVSTP
      101 FFNRLWYRL LSSRFSWLKWS YCPRFELDYL EAFGLLSDFL DHQAVIKFFE
      151 LETHFSYYPV SGFVAPHQYL SLLQDRYFPI ASVMRTLDDK NFSLTPDLIH
      201 DLLGHVPWLL HPSFSEFFIN MGRLPKIVIE KVQALPSKKQ RIQTQSNLI
      15  251 AIVRCFWFTV ESGLIENHEG RKAYGAVLIS SPQELGHAFI DNVRVLPLEL
      301 DQIIRLPFNT STPQETLFSI RHFDELVELT SKLEWMLDQG LLESIPLYNQ
      351 EKYLSGFEVL CQ*
```

The cp7380 nucleotide sequence <SEQ ID 202> is:

```

20      1  GTGCACTACT GCGAGAGAAC CCTGGACCCA AAGTATATTC TGAAGATTGC
      51  TCTAAAGCTG AGACAATCAC TTTCCCTGTT CTTCCAGAAC AGCCAATCAC
      101 TCCAACGTGC ATACTCGACC CCATATTCC TACTACGAAT CATTCTACAA
      151 AAGGAAAATA AAGAGAAGCA AGCTTTAGCT CGACACAAAT GCATTCTAT
      201 TTTAGAATTT TTCAAAAAC TACTCTTTGT TCATCTTCTG TCATTATCAA
      251 AGAATCAAAG GGAAGGTTGC TCCACTGATA TGGCTGTTGT AAGCACTCCC
      25  301 TTTTTTAATC GGAATTTATG GTATCGACTC CTTTCCTCAC GGTTTTCTCT
      351 ATGGAAAAGC TATTGTCCAA GATTTTCTCT TGATTACTTA GAAGCTTTCG
      401 GTCTCTTTTC TGATTCTTAA GACCATCAAG CAGTCATTAA ATTCTTCGAA
      451 TTAGAAACAC ATTTTTCCTA TTATCCCGTT TCAGGATTGT TAGCTCCCCA
      501 TCAATACTTG TCTCTGTTGC AGGACCGTTA CTTTCCCAT GCCTCTGTAA
      30  551 TGCGAACTCT CGATAAAGAT AATTTCCTCT TAACCTCTGA TCTCATCCAT
      601 GACCTTTTAG GGCACGTGCC TTGGCTTCTA CATCCCTCAT TTTCTGAATT
      651 TTTCTATAAC ATGGGAAGAC TCTTCACTAA AGTCATAGAA AAAGTACAAG
      701 CTCCTCCTAG TAAAAACAA CGCATACAAA CCTACAAAG CAATCTGATC
      751 GCTATTGTAC GCTGCTTTTG GTTTACTGTT GAAAGCGGAC TTATTGAAAA
      35  801 CCATGAAGGA AGAAAAGCAT ATGGAGCCGT TCTTATCAGT TCTCCTCAGG
      851 AACTTGGACA CGCTTTCAT TATAACGTAC GTGTTCTCCC TTTAGAATTG
      901 GATCAGATTA TTCGTCTTCC CTTCAATACA TCAACTCCAC AAGAGACTTT
      951 ATTTTCAATA AGACATTTTG ATGAACTGGT AGAACTCACT TCAAAATTAG
      1001 AATGGATGCT CGACCAAGGT CTGTTAGAA CAAATCCCCT TTACAATCAA
      40  1051 GAGAAATATC TTTCTGGTTT TGAGGTACTT TGCCAATGA
```

The PSORT algorithm predicts inner membrane (0.1362).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 101A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 101B) and for FACS analysis (Figure 101C).

These experiments show that cp7380 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 102

The following *C.pneumoniae* protein (PID 4376904) was expressed <SEQ ID 203; cp6904>:

-135-

1 MMNYEDAKLR GQAVAILYQI GAIKFGKHIL ASGEETPLYV DMRLVISSPE  
 51 VLQTVATLIW RLRPSFNSSL LCGVPYTALT LATSISLKYN IPMVLRRKEL  
 101 QNVDPDAIK VEGLFPPGQT CLVINDMVSS GKSLIETAVA LEENGLVVRE  
 151 ALVFLDRRKE ACQPLGPQGI KVSSVFTVPT LIKALIAYGK LSSGDLTLAN  
 201 KISELLEIES \*

The cp6904 nucleotide sequence <SEQ ID 204> is:

1 ATGATGAACT ACGAAGATGC AAAATTACGC GGTCAAGCTG TAGCAATTCT  
 51 ATACCAAATC GGAGCTATAA AGTTCGGAAA ACATATTCTC GCTAGCGGAG  
 101 AAGAACTCC TCTGTATGTA GATATGCGTC TTGTGATCTC CTCTCCAGAA  
 151 GTTCTCCAGA CAGTGGCAAC TCTTATTTGG CGCCTCCGCC CCTCATTCAA  
 201 TAGTAGCTTA CTCTGCGGAG TCCCTTATAC TGCTCTAACC CTAGCAACCT  
 251 CGATCTCTTT AAAATATAAC ATCCCTATGG TATTGCGAAG GAAGGAATTA  
 301 CAGAATGTAG ACCCCTCGGA CGCTATTAAA GTAGAAGGGT TATTACTCC  
 351 AGGACAAACT TGTTTAGTCA TCAATGATAT GGTTCCTCA GGAATCTTA  
 401 TAATAGAGAC AGCAGTCGCA CTGGAAGAAA ATGGTCTGGT AGTTCGTGAA  
 451 GCATTGGTAT TCTTAGATCG TAGAAAAGAA GCGTGTCAAC CACTTGGTCC  
 501 ACAGGGAATA AAAGTCAGTT CGGTATTTAC TGTACCCACT CTGATAAAG  
 551 CTTTGATCGC TTATGGGAAG CTAAGCAGTG GTGATCTAAC CCTGGCAAAC  
 601 AAAATTTCCG AAATCTAGA AATTGAATCT TAA

The PSORT algorithm predicts cytoplasm (0.0358).

The protein was expressed in *E.coli* and purified as a his-tag product (Figure 102A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 102B) and for FACS analysis.

The cp6904 protein was also identified in the 2D-PAGE experiment.

These experiments show that cp6904 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 103

The following *C.pneumoniae* protein (PID 4376964) was expressed <SEQ ID 205; cp6964>:

1 MKKLIALIGI FLVPIKNTN KEHDAHATVL KAARAKYNLF FVQDVFPVHE  
 51 VIEPISPDCL VHYEGWV\*

The cp6964 nucleotide sequence <SEQ ID 206> is:

1 ATGAAAAAAT TGATTGCTTT GATAGGGATA TTTCTTGTTT CAATAAAAGG  
 51 AAATACCAAT AAGGAACAG ACGCTCACGC GACTGTTTAA AAAGCGGCCA  
 101 GAGCAAAGTA TAATTGTGTC TTTGTTTCAG ATGTTTTCCT TGTACACGAA  
 151 GTTATCGAGC CTATTTCTCC CGATTGCCTG GTACATTATG AAGGGTGGGT  
 201 TTGA

The PSORT algorithm predicts inner membrane (0.091).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 103A) and also in his-tagged form. The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (Figure 103B) and for FACS analysis (Figure 103C).

These experiments show that cp6964 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 104

The following *C.pneumoniae* protein (PID 4377387) was expressed <SEQ ID 207; cp7387>:

-136-

```

1  LNFQKIDHNH LYLTCGLDLG VACPILSTDC LPNYSEKASH EVLVYSKFRG
51  ISGEP SRLAT SGNDYYYSIV SLPGLRYEV TSPSGRHDFN IDMHVAPKIG
101 AVLSHGTRBA KEIPGSSKDY AFFSLTARES LMISEKLAMT FQVSEVIQNC
151 YSQCTKVTKT NLKEQYRHLN HNTGFELSVK SAF*

```

5 The cp7387 nucleotide sequence <SEQ ID 208> is:

```

1  TTGAATTTTG CAAAGATTGA TCACAATCAT CTCTACCTTA CATGTTTGGG
51  AGATCTTGGT GTAGCTTGTC CTATACTTTC TACAGATTGT CTACCTAATT
101 ATAGCGAGAA AGCATCTCAT GAGGTTCTTG TTTATAGTAA ATTTAGATGC
151 ATTTCTGGAG AGCCATCTCG ACTTGCAACT TCAGGAAATG ACACATATTA
10  201 TTCTATAGTA AGTTTACCTA TAGGACTCCG TTACGAAAGT ACTTCACCAT
251 CAGGACGTCA TGATTTCAAT ATTGATATGC ATGTAGCTCC AAAGATAGGT
301 GCAGTACTCT CTCATGGAAC ACGAGAGGCT AAAGAGATCC CAGGATCTTC
351 AAAAGACTAT GCATTTTTTA GCTTGACTGC TAGAGAAAGT TTAATGATTT
401 CTGAAAAGCT TCGGATGACT TTCCAAGTTA GCGAAGTTAT TCAGAAATGT
15  451 TATTCACAAT GTACTAAAGT AACGAAAACG AATTTAAAG AACAGTATAG
501 GCACCTATCC CACAATACAG GGTTTGAGTT AAGCGTCAAG TCTGCATTCT
551 AA

```

The PSORT algorithm predicts inner membrane (0.043).

The protein was expressed in *E.coli* and purified as a his-tagged-fusion product (Figure 104A) and also as a GST-fusion (Figure 104B). The recombinant proteins were used to immunise mice, whose sera were used in a Western blot and for FACS analysis (Figure 104C; his-tagged).

These experiments show that cp7387 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

#### Example 105

25 The following *C.pneumoniae* protein (PID 4376281) was expressed <SEQ ID 209; cp6281>:

```

1  MFLQFFHPIV FSDQSLSFLP YLGKSSGIIE KCSNIVEHYL HLGGDTSVII
51  TGVSGATFLS VDHALPISKS EKIIKILSYI LILPLILALF IKIVLRIILF
101 FKYRGLILDV KKEDLKKTLT PDQENLSLPL PSPTTLKKIH ALHILVRSGK
151 TYNELIQEGF SFTKITDLGQ APSPKQDIGF SYNSSLPNFY FHSLSVSPNI
30  201 SGEERALNYH KBQOEEMAVK LRTMQACSFV FRSLLHPSMQ TKDKKAGFGL
251 LTFPPWKIYP L*

```

The cp6281 nucleotide sequence <SEQ ID 210> is:

```

1  ATGTTTCTTC AGTTTFTTCA TCCTATAGTC TTCTCGGATC AGTCCTTATC
51  TTTTCTTCCT TACCTAGGAA AAAGCTCTGG CATTATTGAA AAATGTTCCA
35  101 ATATCGTTGA ACACTATTTA CATTTGGGAG GAGACACTTC TGTTATCATC
151 ACAGGAGTTT CTGGAGCTAC CTTTCTATCT GTTGATCATG CCCTCCCAAT
201 CTCGAAATCT GAAAAAATAA TAAAAATTCT CTCCTATATT TTAATCTTTC
251 CTCTGATTCT AGCTCTCTTT ATTAAGATCG TTTTACGCAT TATCTTATTC
301 TTCAAGTATC GTGGTCTAAT CCTAGATGTT AAGAAGGAGG ATTTGAAAAA
40  351 AACACTTACA CCTGACCAAG AAAACCTCAG TCTTCTTTTA CCATCTCCTA
401 CAACATTAAG GAAAAATCAT GCGCTACACA TTTTAGTGCG TTCTGGAAAA
451 ACCTATAACG AGCTTATACA AGAAGGGTTT TCTTTCACTA AAATCACAGA
501 TCTTGGTCAA GCTCTTTCAC CAAAGCAAGA TATTGGCTTC TCTTATAATT
551 CCTTCTCTCC TAACTTCTAT TTTTCATTCCT TGGTATCTGT TCCAAATATT
45  601 TCAGGCGAGG AACGGGCTCT TAATTATCAT AAAGAACAAC AAGAGGAAAT
651 GGCTGTTAAA TTAATAACAA TGCAAGCGTG TTCTTTTGTC TTCCGATCCC
701 TGCATTTACC TTCAATGCAA ACGAAGGACA AAAAGGCTGG ATTTGGACTA
751 CTGACGTTTT TCCCTTGGA AATCTACCCC CTATAA

```

The PSORT algorithm predicts inner membrane (0.5373).

50 The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 105A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 105B) and for FACS analysis.

These experiments show that cp6281 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

#### Example 106 and Example 107

- 5 The following *C.pneumoniae* protein (PID 4376306) was expressed <SEQ ID 211; cp6306>:

```

1  MGNHETYIHP  GVLPSASHAQD  VSRSTVYPSR  SFIMRRMLMG  WNFNRVPSKS
51  SEQLMDGHRI  PLIFFGKHP  TISILNVNRF  SWLSIFYNGE  RGF*
```

The cp6306 nucleotide sequence <SEQ ID 212> is:

```

1  ATGGGAAACC  ATGAGACCTA  TATACATCCA  GGAGTGCTCC  CGAGTAGTCA
10  51  TGCTCAGGAT  GTTAGCAGAT  CTACAGTTTA  CCCAGTCGA  AGTTTATCA
    101  TGAGACGTAT  GCTCATGGGC  TGGAAATTCA  ATCGTGTTC  CTCGAAGAGC
    151  TCCGAGCAGT  TAATGGATGG  TCATCGCATA  CCTCTTATAT  TTTTGGGAA
    201  GCATCATCCT  ACTATATCTA  TTTTAAATGT  CAATAGATTT  TCTTGGCTCT
    251  CCATTTTTTA  CAATGGAGAA  AGGGGGTTTT  GA
```

- 15 The PSORT algorithm predicts cytoplasm (0.167).

The following *C.pneumoniae* protein (PID 4376434) was also expressed <SEQ ID 213; cp6434>:

```

1  MSESINRSIH  LEASTPFFIK  LTNLCSRLV  KITSLVISLL  ALVGAGVTLV
51  VLFVAGILPL  LPVLILBIL  ITVLVLLFCL  VLEPYLIERP  SKIKELPKVD
101  ELSVETDST  L*
```

- 20 The cp6434 nucleotide sequence <SEQ ID 214> is:

```

1  ATGTCTGAAA  GTATTAACAG  AAGCATTCAT  TTAGAAGCCT  CTACACCATT
51  TTTTATAAAA  TTAACGAATC  TCTGTGAAAG  TAGATTAGTT  AAGATCACTT
25  101  CTCTTGTTAT  TTCTCTATTA  GCTTTAGTGG  GTGCGGGAGT  CACTCTTG TG
    151  GTTTTATTTG  TAGCTGGGAT  CCTTCCTTTA  CTTCTGTAC  TCATCTTAGA
    201  AATTATTTTA  ATAACCGTCC  TTGCTTGCT  TTTTGTTTG  GTATTGGAAC
    251  CTTATTTAAT  AGAAAAACCT  AGTAAATAA  AGGAAGTACC  TAAAGTAGAC
    301  GAGCTATCTG  TAGTAGAAAC  GGACAGTACT  CTTTAA
```

The PSORT algorithm predicts inner membrane (0.6859).

- 30 The proteins were expressed in *E.coli* and purified as his-tag products (Figure 106A; 6306 = lanes 2-4; 6434 = lanes 8-10). The recombinant proteins were used to immunise mice, whose sera were used in Western blots (Figures 106B & 107) and for FACS analysis.

These experiments show that cp6306 & cp6434 are surface-exposed and immunoaccessible proteins, and that they are useful immunogens. These properties are not evident from the sequences alone.

#### Example 108

- 35 The following *C.pneumoniae* protein (PID 4377400) was expressed <SEQ ID 215; cp7400>:

```

1  MRVMRFFCLF  FLGFLGSFHC  VAEDKGVDLF  GVWDDNQITE  CDDSYMTEGR
51  EEVEKVVD A
```

The cp7400 nucleotide sequence <SEQ ID 216> is:

```

40  1  GTGAGAGTTA  TGAGATTTTT  TTGCTATTT  TTTCTGGGT  TCCTAGGATC
    51  TTTTCATTGT  GTTCTGAAG  ACAAGGGCGT  GGATTATTT  GGAGTCTGGG
    101  ACGATAACCA  AATTACAGAG  TGTGACGATA  GTTACATGAC  AGAGGTCGT
    151  GAAGAGGTTG  AAAAGGTAGT  GGACGCTTAG
```

The PSORT algorithm predicts periplasmic space (0.924).



-138-

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 108A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 108B) and for FACS analysis.

These experiments show that cp7400 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 109

The following *C.pneumoniae* protein (PID 4376395) was expressed <SEQ ID 217; cp6395>:

```

1  MENAMSSSFV YNGPSWILKT SVAQEVFKKH GKGIQVLLST SVMFLFIGLGV
51  CAFIFPQYLI VFVLTIALLM LAISLVLFLL IRSVRSSMVD RLWCSEKGYA
101 LHQHENGPFLL DVKRVQQILL RSPYIKVRAL WPSGDIPEDP SQAAVLLLSLSP
151 WTEFFSSVDVE ALLPSPQKKE GKVIDPVLPK LSRIERVSLL VFLSAFTLDD
201 LNEQGVNPLM NNEEFLEFFIN KKAREHGIQD LKHEIMSSLE KTGVPPLDPSM
251 SFQVSQAMFS VYRYLRQRDL TTSELRCFHL LSCFRGDVVH CLASFENPKD
301 LADSDPLEAC KNVEWGEFIS ACEKALLKNP QGISIKDLKQ FLVR*
```

The cp6395 nucleotide sequence <SEQ ID 218> is:

```

1  ATGGAAGATG CTATGTCATC ATCGTTTGTG TATAATGGGC CTCGTGGAT
51  TTTAAAAACG TCAGTAGCTC AGGAGGTATT TAAAAAGCAC GGTAAGGGGA
101 TTCAGGTTCT CTTAAGTACT TCAGTGATGC TTTTATAGG TCTTGGAGTC
151 TGTGCCTTTA TATTTCTCTA ATATCTGATT GTTTTGTGTT TGAATATAGC
201 TTTGCTTATG CTCGCTATAA GCTTGGTATT GTTCTCTTA ATACGTTCTG
251 TACGCTCTTC AATGGTAGAT CGTTTGTGGT GTTCTGAAAA AGGATATGCT
301 CTTTCATCAAC ATGAGAACGG GCCTTTTTTG GATGTGAAGC GTGTACAGCA
351 AATTCTTCTA AGATCACCTT ATATTAAAGT TCGGGCTTTA TGGCCGTCCTG
401 GAGATATCCC TGAGGATCCT TCACAAGCTG CGGTTCTATT ACTTTCTCCT
451 TGGACTTTCT TTTCAATCCG GTATGTAGAG GCTTTATTAC CGAGTCCTCA
501 AGAAAAGGAG GGTAAGTATA TAGATCCTGT GCTGCCTAAG TTGTCTAGGA
551 TAGAGAGAGT CTCACTTTTA GTGTTTGTGA GTGCATTTAC TTTGGATGAC
601 TTAAACGAAC AGGGAGTCAA TCCTTTGATG AATAATGAGG AATTTTATT
651 TTTTATAAAT AAGAAAGCGC GTGAGCATGG GATTTCAGGAT TTAAACACG
701 AGATTATGTC TTCGTTAGAG AAAACAGGAG TGCCATTAGA CCCCTCAATG
751 AGTTTTCGAG TTTTCAACAG GATGTTTCTT GTATATCGCT ACTTGAGACA
801 AAGGGATTTA ACGACTTCAG AATTAAGATG TTTTCACTC TTAAGTTGTT
851 TTAAAGGGGA TGTGGTTCAT TGTTTAGCTT CATTGAAAA CCCTAAAGAT
901 TTAGCAGATT CTGACTTTT AGAAGCTTGT AAGAACGTGG AATGGGGTGA
951 GTTTATTTTC GCATGTGAGA AGGCTCTTTT AAAGAATCCG CAAGGAATTT
1001 CCATTAAAGGA TCTAAACAA TTTTATAGTA GGTAA
```

The PSORT algorithm predicts inner membrane (0.6307).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 109A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 109B) and for FACS analysis.

These experiments show that cp6395 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 110

The following *C.pneumoniae* protein (PID 4376396) was expressed <SEQ ID 219; cp6396>:

```

1  MIEFAFVPHT SVTADRIEDR MACRMNKLST LAITSLCVLI SSVCMIGIL
51  CISGTVGTYA FVVGIIFSVL ALVACVFFLY FFYFSSEKFK CASSQEFRLF
101 PIPAVVSALR SYEYISQDAI NDVIKDTMQL STLSSLLDPE APFLEFPYFN
151 SLIVNHSMEKE ADRLSREAFI ILLGEITWKO CETKILEPWLK DPNITPDDEFW
201 KLLKDHFDLK DFKKRIATWI RKAYPEIRLP KKHCLDKSIY KGCCCKFLLLS
```

-139-

251 ENDVQYQRLH HKVCYFSGEF PAMVLGLGSE VPMVLGLPKV PKDLTWEMFM  
301 ENMPVLLQSK REGHWKISLE DVASL\*

The cp6396 nucleotide sequence <SEQ ID 220> is:

```

5      1  ATGATCGACT TTGCTTTTGT TCCTCATACC TCCGTGACAG CGGATCGGAT
      51  TGAGGATCGC ATGGCCTGTC GCATGAACAA GTTGTCTACT TTAGCAATTA
101    101  CAAGTCTTTG TGTATTGATC AGTTCAGTTT GTATTATGAT TGGGATTTTA
      151  TGCATTTCTG GAACGGTTGG GACCTATGCA TTTGTGTAG GAATTATTTT
201    201  TTCTGTGCTT GCTTTGGTAG CATGTGTTT CTTCTTTTAT TTCTTTTATT
      251  TTTCTTCTGA GGAATTTAAG TGTGCTTCTT CGCAGGAGTT TCGTTTTTTG
10     301  CCTATACCAG CTGTGGTTTC TGCATTGCGT TCCTATGAAT ACATTTCTCA
      351  GGACGCTATC AATGACGTTA TAAAAGATAC GATGCAGTTG TCTACCCTTT
401    401  CTTCTCTTTT AGATCCCGAA GCTTTTTTCT TAGAATTTCC TTATTTTAAC
      451  TCTTTGATAG TGAATCATTC GATGAAGGAA GCGGATCGTT TGTCTCGAGA
15     501  GGCTTTTTTG ATTTTATTAG GTGAGATTAC TTGGAAGGAT TGTGAAACAA
      551  AAGATTTGCG ATGGTTGAAA GATCCTAATA TCACTCCTGA TGATTTCTGG
      601  AAGCTATTAA AAGACCATTT CGATTTAAAG GACTTTAAGA AGAGGATCGC
651    651  CACTTGGATA CGGAAGGCCT ATCCAGAAAT TAGATTACCG AAGAAGCATF
      701  GTTTAGATAA GTCTATCTAT AAGGGGTGTT GTAAGTTTTT ATTACTTTCT
20     751  GAGAATGATG TGCAATATCA GAGGTTATTA CATAAGGTCT GTTATTCTC
      801  TGGGGAGTTT CCTGCCATGG TTTTAGGTTT GGAAGTGAA GTGCCATGG
      851  TGTTAGGACT CCTAAGGTT CCAAGGATC TTACCTGGGA GATGTTTATG
901    901  GAAATATATG CTGTTCTTCT GCAAAGCAA AGAGAGGGGC ATTGAAAAT
951    951  CTCCTTGGA GACGTAGCCT CTCTTTAA

```

The PSORT algorithm predicts inner membrane (0.6095).

- 25 The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 110A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 110B) and for FACS analysis.

These experiments show that cp6396 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### 30 Example 111

The following *C.pneumoniae* protein (PID 4376408) was expressed <SEQ ID 221; cp6408>:

```

35     1  MNTSLKRPLK SHFDVVSFL RPEHLKKTRE SLKEGSISLD QLMQIEDIAI
      51  QDLIKQKAA GLSFTDGEF RRATWHYDFM WGFHGVGHRH ATEGVFFDGE
101    101  RAMIDDTYLT DKISVSHHPF VDHFKPVKAL EDEFTTAKQT LPAPAQFLKQ
      151  MIFPNNIEVT RKFYPTNQEL IEDIVAGYRK VIRDLVDAGC RYLQLDDCTR
201    201  GGLVDPVCS WYGIDEKGLQ DLIQQYLLIN NLVIADRPDD LVVNLHVCRG
      251  NYHSKFFASG SYDFIAKPLF EQTNVDGYL EFDHERSGDF SPLTFISGEK
301    301  TVCLGLVTSK TPTLENKDEV IARIHQAADY LPLERLSLSP QCGFASCEIG
351    351  NKLTEEQWA KVALVKEISE EVWK*

```

40 The cp6408 nucleotide sequence <SEQ ID 222> is:

```

45     1  ATGAATACTT CACTAAAAAG ACCTCTGAAA TCTCATTTTG ATGTTGTCGG
      51  TAGTTTTTTG CGTCCTGAGC ATTTAAAAAA AACTAGAGAA AGCCTTAAAG
101    101  AAGGCTCTAT TTCTCTAGAT CAACTCATGC AAATTGAGGA TATCGCTATC
      151  CAAGATTTGA TCAAAAAACA AAAAGCAGCA GGTCTTTCTT TTATTACTGA
201    201  TGGAGAATTC CGCAGAGCTA CGTGGCATTG CGACTTCATG TGGGGTTTTT
      251  ATGGCGTAGG TCACCACAGA GCTACAGAAG GAGTTTCTTT TGATGGAGAA
301    301  CGCGCTATGA TCGATGATAC CTATCTGACA GACAAGATCT CTGTATCTCA
      351  CCACCCATTT GTGGATCACT TTAAATTTGT AAAAGCTCTA GAAGATGAAT
50     401  TTACGACTGC AAAGCAAACT CTTCTGCAC CGGCACAGTT TTTAAAGCAG
      451  ATGATCTTCC CTAATAATAT AGAGGTCACA CGTAAATTCT ATCCTACAAA
      501  TCAGGAGCTA ATTGAAGATA TTGTTGCAGG TTATCGTAAA GTCATTCGCG
      551  ATCTTTATGA TGCTGGCTGC CGCTATCTCC AATTAGATGA CTGTACTCGG
601    601  GGAGGTTTAG TAGACCTCG AGTCTGTTTC TGGTATGGTA TCGATGAAAA
      651  AGGTCTTCAA GATCTGATTC AACAATATCT TCTGATTAAT AATCTTGTA
55     701  TTGCAGATCG TCCCGATGAT CTAGTCGTTA ATTTACATGT ATGCCGTGGG

```

-140-

```

751  AACTACCACT  CAAAATTCTT  TGCTAGTGGT  AGTTATGACT  TTATTGCAAA
801  GCCCCTATTC  GAACAAACAA  ATGTAGACGG  CTACTATTTA  GAGTTTGATC
851  ATGAGCGTTC  TGGAGACTTC  TCTCCTCTCA  CCTTCATTTC  TGGAGAAAAA
901  ACTGTCTGCT  TAGGTCTTGT  TACCAGCAAA  ACCCCTACAC  TTGAAAATAA
5    951  GGATGAGGTC  ATTGCTCGCA  TACATCAAGC  AGCAGACTAC  CTGCCCTTGG
1001 AAAGACTCTC  TCTAAGTCCA  CAGTGTGGTT  TTGCTTCATG  TGAAATAGGA
1051 AATAAATTAA  CAGAAGAAGA  GCAATGGGCT  AAAGTTGCTC  TAGTAAAGA
1101 AATTTCGGAA  GAAGTTTGA  AATAA

```

The PSORT algorithm predicts cytoplasm (0.2171).

- 10 The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 111A) and also as a his-tagged product. The his-tag protein was used to immunise mice, whose sera were used in a Western blot (Figure 111B) and for FACS analysis.

These experiments show that cp6408 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### 15 Example 112

The following *C.pneumoniae* protein (PID 4376430) was expressed <SEQ ID 223; cp6430>:

```

1  MKLYSISSDV  DTPWIFQLMS  KVDSYFLGG  NRIKVVSIWM  QEPNLIIGKV
51  ENVRISTIVK  ILKILSFLIF  PLILIALALH  YFLHAKYANH  LLVSKILERA
101 PQYVIPGRS  GDTASHYKLT  TLVPVSQKNL  QAMGSPNPLEV  EAALRTTKPS
20  151  FFCVPAKYRQ  IIISSHGIRF  SLDLBQLADD  INLDSVSWPT  EYLNSTMDFC
201 SKADKRVIQN  VQNLRTGTYY  NSVGKRSLLK  FMLQHLFIDG  ITQENPEALP
251 NNTSGRLTLF  PSVRYIYSHF  TPQNPTIWPQ  VFFRQGFLDE  DRGGGFLEILE
301 QLQELGVRFP  ICPSQGPDPN  NFQGFQGIRI  YWEDSYQPNK  EV*

```

The cp6430 nucleotide sequence <SEQ ID 224> is:

```

25 1  ATGAAACTTT  ATAGCATCTC  TTCAGATGTA  GATACACCTT  GGATATTTCA
51  GCTTATGTCA  AAGGTAGATT  CTTATCTTTT  CTTAGGCGGG  AATAGAATCA
101 AGGTTGTATC  TATAGTTATG  CAAGAACCTA  ACTTAATTAT  TGGAAAAGTA
151 GAAAACGTTT  GGATCTCCAC  AATAGTGAAG  ATATTAAAGA  TTTTATCCTT
201 CTTAATCTTC  CCTCTGATTT  TAATCGCTTT  AGCCCTACAC  TATTTTCTAC
30  251  ATGCTAAATA  TGCTAATCAC  TTAATGTAT  CTAAGATTTT  AGAAAGAGCT
301 CCTCAGTATG  TGCCTATTCC  TGGTCGTTCA  GGAGACACGG  CGTCTCATT
351 TAAATTAACA  ACATTGGTTC  CAGTATCCCA  AAAAAATCTA  CAAGCTATGG
401 GATCAAATCC  TCTAGAAGTT  GAAGCGGCTC  TTCGAACTAC  AAAACCCTCT
451 TTTTCTGTG  TACCTGCAAA  ATACCGTCAG  ATTATAATTT  CAAGTCACGG
35  501  AGGTCGCTTT  TCTTTAGATC  TTGAACAAC  TGCTGATGAC  ATTAATTTAG
551 ATTCGGTTTC  CTGGCCTACG  GAGTATCTTA  ACTCTACTAT  GGATTTTTCG
601 AGCAAGGCAG  ATAAACGTGT  TATACAGAAT  GTACAAAATC  TCGGACACGG
651 AACTTACATA  AATTCTGTAG  GAAAGCGTAG  CCTTTTAAAA  TTCATGTTAC
701 AGCACCTATT  TATTGATGGG  ATCACACAAG  AAAACCCTGA  AGCCCTTCCT
40  751  AACCAATACAT  CTGGAAGACT  GACTCTATTC  CCTAGTGTTC  GTTATATCTA
801 TTCTCATTTT  ACTCCACAAA  ATCCTACAAT  ATGGCCGCAA  GTCTTTTTC
851 GACAAGGTCC  TCTAGATGAA  GATCGAGGAG  GAGGATTTGA  GATCTTAGAG
901 CAATTACAAG  AGTTAGGAGT  TAGGTTTCCA  ATTTGCCCTT  CTCAGGACC
951 AGACAATCCT  AATTTTCAAG  GTTTTCAAGG  GATTCGTATC  TATTGGGAAG
45 1001  ATTCCTATCA  ACCCAATAAG  GAGGTTTAA

```

The PSORT algorithm predicts inner membrane (0.5140).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 112A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 112B) and for FACS analysis.

- 50 These experiments show that cp6430 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

**Example 113**

The following *C.pneumoniae* protein (PID 4376439) was expressed <SEQ ID 225; cp6439>:

```

1  MSYDTLFLKNL EKEDSVHKIC NEIFALVPRL NTLACTEAI KNLPKADIVH
51  HLPGTITPQL AWILGVKNGF LKWSYNSWTN HRLSPKNPH KQYSNIFRNF
5  101  QDICHEKDPD LSVLQYNILN YDFNSFDRVM ATVQHRFPF GGIQNEEDLL
151  LIFNNYLQQC LDDTIVYTEV QQNIRLAHLV YPSLPEKHAR MKFYQILYRA
201  SQTFSKHGIT LRFLNCFNKT FAPQINTQEP AQEAVQWLQE VDSTFPGLFV
251  GIQSAGSESA PGACPKRLAS GYRNAYDSGF GCEAHAGEGI ETRTIFSSAK
301  VNPEGLIEIT RVTFFSLKRR QPSSLPPIRV CQLG*

```

10 The cp6439 nucleotide sequence <SEQ ID 226> is:

```

1  ATGTCTTATG ATACGTTATT CAAGAATCTT GAAAAGGAAG ATTCTGTACA
51  TAAGATATGC AATGAGATCT TTGCATTAGT ACCACGACTC AATACAATCG
101  CTTGCACCGA AGCTATCATC AAAAACCTCC CCAAAGCAGA TATCCATGTA
151  CACCTTCCTG GGACCATAAC ACCTCAATTA GCTTGGATTT TAGGTGTGAA
15  201  AAATGGGTTC TTAAATGGT CTTATAATT CTTGGACCAAT CATCGATTAC
251  TTTCTCCTAA GAATCCTCAT AAACAATACT CCAATATTTT CCGAAACTTT
301  CAAGATATCT GTCACGAAAA GGATCCGGAT TTAAGTGAT TACAATATAA
351  TATCTTAAAT TACGATTTTA ATAGCTTTGA TAGAGTGATG GCTACAGTAC
401  AAGGACATCG CTTTCCTCCT GGAGGAATCC AAAATGAAGA AGACCTTCTT
20  451  CTCATTTTCA ATAACATATC CCAGCAATGT CTGGACGATA CTATCGTGTA
501  TACTGAAGTA CAACAAAATA TCCGCCTTGC CCATGTTTTG TATCCTTCAT
551  TACCTGAAAA GCACGCGCGT ATGAAGTTT ATCAAATCTT GTATCGTGCT
601  TCGCAAAACG TTTCAAAACA CGGGATTACT TTACGATTTT TAAACTGCTT
651  CAATAAAACA TTTGCTCCAC AAATAAACAC ACAAGAACCT GCCCAAGAAG
25  701  CTGTTCATG GCTCCAAGAG GTTGATTCTA CATTTCTCTG TCTATTTGTA
751  GGGATACAAT CCGCAGGATC AGAATCTGCG CCCGAGCCT GTCTAAGCG
801  ATTAGCTTCT GGATATAGAA ATGCTTATGA CTCAGGGTTT GGTGTGAAG
851  CTCATGCTGG AGAAGGCATA GAGACCCGGA CTATTTTTC GTCAGCTAAG
901  GTAAATCCAG AGGGATGTAT CGAGATAACC CGAGTGACTT TCTCGTCTCT
30  951  TAAACGAAAA CAGCCATCTA GTTTACCCAT AAGAGTTACT TGCCAGTTAG
1001 GATAA

```

The PSORT algorithm predicts cytoplasm (0.1628).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 113A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 113B) and for FACS analysis.

These experiments show that cp6439 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

**Example 114**

The following *C.pneumoniae* protein (PID 4376440) was expressed <SEQ ID 227; cp6440>:

```

40 1  LQSARRHLNT IFILDFGSQY TYVLAKQVRK LFVYCEVLPW NISVQCLKER
51  APLGIILSGG PHSVYENKAP HLDPEIYKIG IPILAICYGM QLMARDFGGT
101  VSPGVGEFGY TPIHLYPCEL FKHIVDCESL DFEIRMSHRD HVTTIPEGFN
151  VIASTSQCSI SGIENTKQRL YGLQFHPEVS DSTPTGNKIL ETFVQEICSA
45  201  PTLWNPLYIQ QDLVSKIQDT VIEVFDEVAQ SLDVQWLAQG TIYSDVIESS
251  RSGHASEVIK SHHNVGGLPK NLKLLVEPL RYLFKDEVRI LGEALGLSSY
301  LLDRHPFPGP GLTIRVIGEI LPEYLALRR ADLIFIEELR KAKLYDKISQ
351  AFALFLPIKS VSVKGDRCYS GTTIALRAVE STDFMTGRWA YLPCDVLSSC
401  SSRIINEIPE VSRVVDISD KPPATIEWE*

```

The cp6440 nucleotide sequence <SEQ ID 228> is:

```

50 1  TTGCAGAGTG CAAGGAGACA TTTGAACACC ATATTTATTC TAGATTTTGG
51  ATCTCAATAT ACTTATGTAT TAGCAAAGCA AGTCCGGAAG TTATTTGTAT
101  ATTGCCAAGT TCTTCCCTGG AATATCTCTG TGCAATGTTT AAAAGAAAGA
151  GCGCCTTTGG GGATCATTCT CTCAGGAGGT CCTCACTCTG TCTATGAAAA

```

-142-

201 CAAGGCTCCA CATTTAGATC CTGAAATCTA TAAACTTGGC ATTCCAATTC  
 251 TAGCTATTTC CTATGGCATG CAGCTTATGG CTAGAGATTT TGGAGGGACT  
 301 GTAAGCCCTG GTGTAGGAGA ATTTGGATAT ACGCCCATCC ATCTGTATCC  
 351 TTGTGAGCTC TTCAAACACA TCGTCGACTG CGAATCTCTA GACACAGAGA  
 401 TTCGGATGAG CCATCGGGAT CATGTTACGA CAATTCCTGA AGGATTTAAT  
 451 GTAAATCGCAT CCACCTCACA ATGCTCGATC TCAGGAATAG AAAATACCAA  
 501 ACAACGGTTG TACGGGCTGC AATTTTCATCC CGAGGTTTCT GACTCCACTC  
 551 CAACGGGAAA TAAGATTCTA GAACTTTTGG TTCAAGAGAT CTGTTCTGCT  
 601 CCCACACTAT GGAATCCCTT GTATATTCAG CAAGACCTTG TAAGTAAAAAT  
 651 TCAAGATACC GTTATTGAAG TATTTGATGA AGTCGCTCAG TCATTAGACG  
 701 TACAATGGTT AGCTCAAGGA ACCATCTACT CAGATGTTAT TGAGTCCTCA  
 751 CGCTCTGGAC ATGCTCCCGA AGTAATAAAA TCACATCATA ATGTAGGGGG  
 801 GCTTCCAAA AATCTTAAGC TGAAGTTAGT CGAGCCCTTA CGTTATTTAT  
 851 TTAAAGATGA AGTTCGAATT TTAGGAGAAG CCCTAGGACT TTCTAGCTAT  
 901 CTCTTGGACA GGCATCCTTT TCCTGGACCT GGCTTGACAA TTCGTGTGAT  
 951 TGGAGAGATC CTTCTGAAT ATCTAGCCAT TTTACGACGG GCGGACCTCA  
 1001 TCTTTATAGA AGAGCTTAGG AAAGCAAAAC TCTACGATAA AATAAGCCAA  
 1051 GCCTTTGCTC TATTTCTTCC TATAAAATCA GTATCTGTAA AAGGAGATTG  
 1101 TAGAAGCTAT GGTATATACCA TAGCATTACG TGCTGTAGAA TCTACAGATT  
 1151 TCATGACAGG ACGATGGGCC TACCTTCCAT GCGATGTTCT CAGTTCTTGC  
 1201 TCATCGCGAA TTATTAATGA AATACCCGAG GTAAGCCGAG TGGTCTATGA  
 1251 TATTTCTGAC AAGCCACCAG CAACTATAGA ATGGAATAG

The PSORT algorithm predicts cytoplasm (0.0481).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 114A) and also as  
 a his-tagged product. The recombinant proteins were used to immunise mice, whose sera were used  
 in a Western blot (Figure 114B) and for FACS analysis.

These experiments show that cp6440 is a surface-exposed and immunoaccessible protein, and that it  
 is a useful immunogen. These properties are not evident from the sequence alone.

### Example 115

The following *C.pneumoniae* protein (PID 4376475) was expressed <SEQ ID 229; cp6475>:

1 MNTYTFSPITL QKSFSLFLLR KLDSYFFFGG TRTQILVITP TNIRLAAKKR  
 51 GCKVSTIEKI IKILSFILLP LVIIAIFILRY FLHKKFDKQF LCIPKVISNE  
 101 DEALLGSRPQ AVEKAVREIS PAFFSIPRKY QLIRIDTPKD DAPSILFPIG  
 151 IEIILKDLCI DTLKQSNLFL KREMDFLGHP BEKALFDSIC SIEKDQEWMS  
 201 LESKKLLITH FLKYLFSVGI BQLNPGFNPE NGRGYFSEIS TAKIHFHQHG  
 251 RYGPISRSSGP IMKEI\*

The cp6475 nucleotide sequence <SEQ ID 230> is:

1 ATGAATACCT ATACCTTCTC TCCTACACTT CAGAAAAGCT TCAGCCTATT  
 51 TCTTTTAGAA AAATTAGACT CTTACTTTTT CTTTGGAGGG ACTCGTACAC  
 101 AAATCTTAGT CATCACACCA ACCAATATTA GATTAGCAGC TAAAAAAGA  
 151 GGGTGAAGG TTTCTACTAT AGAAAAGATA ATCAAGATCC TCTCTTTTAT  
 201 CCTGCTGCCC CTAGTTATCA TTGCCTTTAT ACTTCGCTAT TTCTTACATA  
 251 AGAAATTCGA TAAACAGTTC TTGTGTATCC CAAAAGTCAT TTCTAACGAA  
 301 GACGAAGCTC TTCTTGGATC TAGACCACAA GCAGTTGAAA AAGCAGTTCG  
 351 AGAAATATCT CCAGCTTCTT TCTCTATACC AAGAAAATAC CAACTTATTA  
 401 GAATCGACAC TCCTAAAGAT GACGCTCCCT CAATCCTTTT CCCTATAGGC  
 451 ATAGAGATCA TTCTCAAAGA TTTATGTATT GATACACTCA AGCAATCTAA  
 501 TCTTTTCCCTT AAAAGAGAAA TGGATTCTCT AGGTCAATCCA GAAGAAAAAG  
 551 CATTAATTCGA CTCGATATGT TCTATAGAAA AAGATCAAGA ATGGATGAGC  
 601 TTGGAAAGTA AAAAAGTTTT AATCAGGCAC TTCCTAAAGT ATCTCTTTGT  
 651 CTCTGGAATC GAACAACTAA ATCCAGGCTT TAACCCAGAG AATGGGCGTG  
 701 GGTAATTTTC AGAAATAAGT ACAGCAAAGA TCCATTTTCA TCAGCACGGT  
 751 CGATATGGGC CAATCCGTTC TTCGGGACCC ATCATGAAGG AAATATAA

The PSORT algorithm predicts inner membrane (0.5373).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 115A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 115B) and for FACS analysis.

These experiments show that cp6475 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

#### Example 116

The following *C.pneumoniae* protein (PID 4376482) was expressed <SEQ ID 231; cp6482>:

```

1  MLVLEALKR EFAHLKDQKP TSDQEITSLY QCLDHLEFVL LGLGQDKFLK
51  ATEDVDLFE SQKAIIDAWNA LLTKARDVLG LGDIGAIYQT IEFLGAYLSK
101 VNRRAFCLAS EIHFLLKTAIR DLNAYYLLDF RWPLCKIEEF VDWGNDCEVEI
151 AKRKLCTFEK ETKEKNESLL REKHAMEKCS IQDLQRKLSL IIEIHDVSL
201 PCFSTPSQE EYQKDCLYQS RLRYLLLLYE YTLCKTSTD FQEQARAKEE
251 FIREKPSLLE LEKGIKQTKK LEFALAKSKL ERGCLVMRKY EAAAKHSLDS
301 MFEETVVKSP RKDTE*
```

The cp6482 nucleotide sequence <SEQ ID 232> is:

```

1  ATGCTAGTAG AGTTAGAGGC TCTTAAAAGA GAGTTGCGC ATTTAAAAGA
51  CCAGAAGCCG ACAAGTGACC AAGAGATCAC TTCACTTTAT CAATGTTTGG
101 ATCATCTTGA ATTCTTTTAA CTCGGGCTGG GCCAGGACAA ATTTTAAAG
151 GCTACGGAAG ATGAAGATGT GCTTTTGTAG TCTCAAAAAG CAATCGATGC
201 GTGGAATGCT TTATTGACAA AAGCCAGAGA TGTTTTAGGT CTGGGGACA
251 TAGGTGCTAT CTATCAGACT ATAGAATTCT TGGGTGCCTA TTTATCAAAA
301 GTGAATCGGA GGGCTTTTGT TATTGCTTCG GAGATACATT TTCTAAAAAC
351 AGCAATCCGA GATTGGAATG CATATTACCT GTTAGATTTT AGATGGCCTC
401 TTTGCAAGAT AGAAGAGTTT GTGGATTGGG GGAATGATTG TGTGAAATA
451 GCAAAGAGGA AGCTATGCAC TTTTGAAAAA GAAACCAAGG AGCTCAATGA
501 GAGCCTTCTT AGAGAGGAGC ATGCGATGGA GAAATGCTCG ATTCAAGATC
551 TGCAAGGAA ACTTAGCGAC ATTATTATTG AATTGCATGA TGTTCCTCTT
601 TTTTGTTTT CTAAGACTCC CAGTCAAGAG GAGTATCAAA AGGATTGTTT
651 GTATCAATCA CGATTGAGGT ACTTATTGTT GCTGTATGAG TATACATGTT
701 TATGTAAGAC ATCCACAGAT TTCAAGAGC AGGCTAGGGC TAAAGAGGAG
751 TTCATTAGGG AGAAATTCAG CCTTCTAGAG CTCGAAAAGG GAATAAACA
801 AACTAAAGAG CTTGAGTTTG CAATTGCTAA AAGTAAGTTA GAACGGGGCT
851 GTTTAGTTAT GAGGAAGTAT GAAGCTGCCG CTAACATAG TTTAGATTCT
901 ATGTTCTGAAG AAGAACTGT GAAGTCGCCG CGGAAAGACA CAGAATAA
```

The PSORT algorithm predicts cytoplasm (0.4607).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 116A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 116B) and for FACS analysis.

These experiments show that cp6482 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

#### Example 117

The following *C.pneumoniae* protein (PID 4376486) was expressed <SEQ ID 233; cp6486>:

```

1  VVVVALFILG IFFLSGSLAF LVHTSCGVLL GAALPILCIG LVLLAVALIV
51  FLCHKHKTRQ DLDYYDQDL D SLVIHKKEIP NDISELRVTF EKLQNLFPFH
101 TKDFSDLSQE LQKFLNCME KWLTLDEVT KFLIVRDRFL ETRRNFTTFG
151 EQVKGIQSNI FDLHEEKSSL YLELYRLRKD LQVLLNFFLL PPGILKVDYD
201 EIEAIKGLFI RLTSRLDKLD VKAQERKKFI NEMSREFKEV EKAFFDIVDRA
251 TTKKLMRAKK EPARLFMGR TESLLEMKN EALKNQGLD PENLSHPELF
301 SPYQQLILN YLNSEIVLHH YEFLISGTVT SGLTLEECEN RMRAASTGLN
```

-144-

351 ALLVRKQLQFR GAIKSAYFEK LTRIEKELRS LQDVIKSLEL ELIHKIKDIV  
401 TEET\*

The cp6486 nucleotide sequence <SEQ ID 234> is:

```

5      1  GTGGTGGTTG TCGCTTTATF TATCCTTGGG ATTTCTTTT TATCTGGTTC
      51  TCTTGCATTC CTTGTTCATA CGTCTTGC GG AGTTCTTTTA GGAGCGGCGC
     101  TTCCCATACT TTGCATAGGT CTTGTTTTAT TGGCTGTAGC TCTTATTGTT
     151  TTCTTATGTC ACAAACACAA GACTCGTCAA GATTAGATT ATTATGATCA
     201  AGATTTAGAT TCTTTGGTGA TTCATAAGAA AGAGATCCCC AATGACATCT
     251  CTGAGTTGCG GGTAAACATTT GAAAAGTTGC AAAATCTGTT TCAGTTCCAT
    301  ACGAAAAGATT TCTCTGATCT AAGCCAAGAG CTTCAGGGTA AATTTATCAA
    351  TTGCATGGAG AAATGGCTAA CTTTGAAGA CGAAGTGACT AAATTTCTTA
    401  TTGTTCGAGA TAGATTTTTA GAAACCAGAA GAAATTTTAC CACTTTTGGA
    451  GAACAGGTTA AAGGGATCCA AAGCAATATT TTTGATTTCG ATGAGGAAAA
    501  GTCCTTCATTA TATTTAGAA TGTATAGGCT TAGGAAAGAC CTCCAAGTTC
    551  TATTAAATTT TTTTCTGCTC CCCCAGGTA TACTCAAGGT AGATTATGAT
    601  GAAATTGAGG CTATCAAAGG TCTGTTTATA AGATTAACTT CTAGATTAGA
    651  TAAGCTTGAT GTGAAAGCTC AGGAACGTAA GAAGTTCATT AATGAAATGA
    701  GTAGGGAATT TAAAGAAGTA GAGAAAGCTT TTGATATTGT CGATAGGGCA
    751  ACAAAGAAAGC TTATGGATAG AGCCAAGAAA GAAAGTCCGG CACGCTTTTT
    801  CATGGGTAGA ACTGAGTCTC TCTTAGAAAT GAAAAAAAT GAAGAAGCCC
    851  TTAAAAATCA GGGGCTAGAT CCTGAAAATC TTTCCCATCC TGAACTTTTT
    901  AGTCCGTATC AACAGCTTTT AATTTTGAAT TATTTAAATA GCGAAATAGT
    951  TCTGCATCAT TATGAGTTCC TTATTCTGG AACAGTAAC TCTGGCCTAA
   1001  CTCTTGAAGA ATGTGAAAT CGAATGAGGG CGGCTTCTAC TGGGTTGAAC
   1051  GCCCTTCTGG TCGTAAAGCT CCAGTTCAGA GGTGCTATAA AATCTGCGTA
   1101  TTTTGAAAAA CTCACAGAGA TTGAAAAAGA GTTACGATCA CTTCAAGACG
   1151  TAATAAAGTC ATTGGAAC TAACATGATCC ATAAGATAAA AGATATAGTG
   1201  ACAGAAGAAA CTTAG

```

The PSORT algorithm predicts inner membrane (0.7474).

- 30 The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 117A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 117B) and for FACS analysis.

These experiments show that cp6486 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### 35 Example 118

The following *C.pneumoniae* protein (PID 4376526) was expressed <SEQ ID 235; cp6526>:

```

40      1  MSPFKKIVNR LLCYISFQKE SRTLPIIIRE PRMTTKSLGS FNSVISKNKI
      51  HFISLGC SRN LVDSEVMLGI LLKAGVESTN BIEDADYLIL NTC AFLKSAR
     101  DEAKDYLDHL IDVKKENAKI IVTGCMTSNH KDELKPWMSH IHYLLGSGDV
     151  ENILSAIBSR ESSEKISAKS YIEMGEVPRQ LSTPKHYAYL KVAEGCRKRC
     201  AFCIIPSIKG KLRSKPLDQI LKEFRILVNK SVKEIILIAQ DLGDYGDLS
     251  TDRSSQLESL LHELLKEPGD YWLRMLYLYP DEVSDGIIDL MQSNPKLLPY
     301  VDIPLQHIND RILKQMRRTT SREQILGFLE KLRKVPOVY IRSSVIVGFP
     351  GETQEEFQEL ADFIGEGWID NLGIFLYSQE ANTPAABELPD QIPEKVRESR
     401  LKILSIIQKR NVDKHNQKLI GEKIEAVIDN YHPETNLLLT ARFYGQAPFV
     451  DPCIIVNEAK LVSHFGERCF IBITGTAGYD LVGRVVKKSQ NQALLKTSKA
     501  *

```

The cp6526 nucleotide sequence <SEQ ID 236> is:

```

50      1  ATGAGTCTTT TTAAGAAAAAT AGTAAATCGC TTACTATGCT ATATTTCTTT
      51  TCAAAAAGAA TCAAGAACTC TCCAATCAT TATTAGAGAA CCTAGGATGA
     101  CAACAAAAG TTTAGGATCT TTCAATTCAG TTATTTCCAA AAATAAAATT
     151  CATTTTATTA GTTTGGGATG CTCTCGGAAC CTGTAGATA GCGAAGTCAT
     201  GCTAGGCATT CTCTTAAGG CAGGTTACGA GTCTACTAAT GAAATTGAAG
     251  ATGCTGACTA TTTAATTTTA AATACCTGTG CGTTTTTAAA AAGTGCTAGA
     301  GATGAAGCTA AAGATTATCT AGACCATCTA ATTGATGTAA AAAAAGAGAA

```

-145-

```

351 CGCTAAAATT ATTGTAAGTG GATGCATGAC TTCCAACCAC AAAGATGAGC
401 TTAAACCCCTG GATGTCACAC ATCCATTACC TACTAGGTTT TGGGGATGTT
451 GAGAATATTC TTTCTGCTAT TGAGTCTCGT GAATCTGGAG AAAAAATCTC
501 TGCAAAGAGT TACATTGAGA TGGGAGAAGT TCCAAGACAG CTTTCCACAC
551 CAAAACACTA TGCCATTTTA AAAGTTGCTG AGGGCTGTAG AAAACGTTGT
601 GCTTTTGTGA TTATTCCTTC CATTAAAGGA AAGCTCCGCA GCAAACCTCT
651 GGATCAAATT CTTAAAGAAT TCCGCATCCT TGTAAACAAG AGTGTGAAAG
701 AGATTATATT GATAGCTCAA GACCTAGGAG ATTATGGAAA GGATCTCTCT
751 ACAGACCGCA GTTCGCAGCT AGAATCACTA TTACATGAGT TACTGAAAGA
801 GCCTGGTGAT TATTGGCTGC GGATGTGTGA TTTATATCCT GATGAAGTGA
851 GTGATGGCAT TATAGATCTT ATGCAATCTA ATCCCAAAC TCTTCCCTAT
901 GTAGATATTC CCTTACAGCA CATTACGAC CGTATTTTAA AGCAAAATGCG
951 AAGAACGACT TCTAGGGAGC AAATCCTAGG ATTCCTAGAA AAATTACGTG
1001 CCAAGGTTCC TCAGGTCTAT ATCCGTCTCT CTGTTATTGT GGGTTTCCCC
1051 GGTGAAACTC AGGAAGAATT CCAGGAGTTA GCTGATTTTA TTGGTGAGGG
1101 TTGGATTGAT AATCTCGGAA TTTCTGTGTA CTCTCAAGAA GCGAATACCC
1151 CGGCAGCAGA ACTCCCTGAC CAGATACCAG AAAAAGTTAA AGAATCGAGG
1201 TTGAAAATTC TATCTCAAAT TCAGAAACGC AATGTGGATA AACATAATCA
1251 GAAGCTCATT GGGGAAAAAA TAGAAGCAGT TATTGATAAC TATCATCCTG
1301 AAACGAATCT TTTACTCACT GCAAGGTTCT ATGGACAAGC TCCTGAAGTG
1351 GACCCTTGTA TTATTGTAAA TGAGGCGAAG CTGTGTTCTC ATTTTGGAGA
1401 AAGATGCTTT ATAGAAATCA CAGGGACTGC TGGTTACGAC CTTGTAGGGC
1451 GTGTTGTAAA AAAATCTCAG AACCAAGCTT TGCTAAAAAC TAGCAAAGCT
1501 TAG

```

The PSORT algorithm predicts cytoplasm (0.1296).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 118A) and also as a his-tagged product. The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (Figure 118B) and for FACS analysis.

These experiments show that cp6526 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

#### Example 119

The following *C.pneumoniae* protein (PID 4376528) was expressed <SEQ ID 237; cp6528>:

```

1 MKNNINNNEC YFKLDSTVDG DLLAANLKTf DTQAQGISST ETFSVQGNAT
51 FKDQVSATGL TSGTTYNLNA QNFTSSQISI DFKNNRLSNC ALPKEDCDPV
101 PANYVRSPEY PFC SKPLIGD FDFNSGESYL PLTGSEYTLY QSRNVNSIFR
151 FIGWKQSTRE LTVGGNTAIQ FLAAGTYIVS FTVGKRWGNW NGWGGAIYIN
201 NGLGQVQCES TIYSGGYAT IGTGTSIYR ASVDVAPNPN DPNASDRYRA
251 GIFYLSNGGS SAGIGNYSFS LLYYPDDR*

```

The cp6528 nucleotide sequence <SEQ ID 238> is:

```

1 ATGAAAAACA ATATTAATAA TAATGAGTGC TATTTTAAAT TAGACTCAAC
51 TGTAGATGGT GATTTGTAG CAGCCAATCT CAAGACCTTT GATACACAGG
101 CCCAAGGAAT CTCATCGACT GAAACATTTT CTGTTCAAGG GAATGCAACA
151 TTFAAGATC AAGTTTCAGC AACTGGATTA ACTTCAGGAA CTACTTATAA
201 TTFAAATGCA CAAAACTTTA CTTCCTCCCA AATCTCTATA GATTTTAAAA
251 ATAATCGTCT GAGTAATTGT GCATTGCCAA AAGAAGACTG CGATCCGGTG
301 CCAGCGAATT ATGTTCTGTT TCCCGAATAT TTTTCTGTT CCAAGCCTCT
351 GATCGGAGAT TTTGATTTTA ACTCAGGGGA ATCTTATTTG CCTCTGACTG
401 GTTCGGAATA TACTCTATAT CAGTCACGTA ATGTAAATAG TATATTTCTG
451 TTTATAGGAT GGAAGCAAAG TACACGAGAA TTAAGTGTAG GGGGAAATAC
501 TGCGATACAA TTTCTTGCA GAGGAACCTA TATCGTTTCA TTTACTGTTG
551 GTAAACGGTG GGGATGGAAT AATGGTTGGG GAGGAGCCAT TTATATCAAT
601 AATGGTTTAG GACAAGTCCA ATGTGAAAGC ACGATTTATA GTGGTGAGG
651 GTATGCAACA ATAGGTACAC TGGGGACCTC AATATATAGA GCCTCTGTAG
701 ATGTAGCTCC TAATCCTAAT GATCCGAATG CTTCGGATCG CTATAGAGCG
751 GGTATTTTCT ATCTCAGTAA CGGTGGTTCT AGTGCAGGTA TAGGGAATTA
801 CTCCTTTTCT CTCTCTATT ATCCGGACGA TAGAGGGTAG

```



The PSORT algorithm predicts cytoplasm (0.1668).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 119A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 119B) and for FACS analysis.

- 5 These experiments show that cp6528 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 120

The following *C.pneumoniae* protein (PID 4376627) was expressed <SEQ ID 239; cp6627>:

```

10      1  MKCSPLTLVP  HIFLKNDCBC  HRSCSLKIRT  IARLILGLVL  ALVSALSFVF
      51  LAAPISYAIG  GTLALAAIVI  LIITLVVALL  AKSKVLPIPN  ELQKIYNRY
      101  PKEVYFVKT  HSLTVNELKI  FINCWKSGTD  LPPNLHKKAE  AFGIDILKSI
      151  DLTLFPEFEB  ILLQNCPLYW  LSHFIDKTES  VAGEIGLNKT  QKVYGLLGPL
      201  AFHKGYYTIF  HSYTRPLLTL  ISESQYKFLY  SKASKNQWDS  PSVKRTCEBI
      251  FKELPHNMIF  RKDVQGISQF  LFLFPESHGIT  WEQAQMIQLI  NPDNWKMLCQ
      15  301  FDKAGGHCSM  ATFGGFLNTE  TNMFPVSSN  YEPTVNFMTW  KELKVILLEKV
      351  KESPMHPASA  LVQKICVNTT  HHQNLKRWQ  FVRNTSSQWT  SSLPQYAFHA
      401  QTYKLEKKIE  SSLPIRSSL*
```

The cp6627 nucleotide sequence <SEQ ID 240> is:

```

20      1  ATGAAGTGTA  GTCCTTTAAC  ACTAGTTCCC  CATATATTTT  TAAAAAATGA
      51  CTGCGAATGT  CATAGATCTT  GTTCTTTAAA  AATTAGGACA  ATTGCCCGAC
      101  TCATTCTTGG  GCTTGTCTTA  GCTCTGTGTA  GCGCACTTTC  TTTTGTPTTC
      151  CTTGCTGCGC  CGATTAGCTA  TGCTATGGA  GGAACTTTAG  CTTTAGCCGC
      201  TATCGTAATC  TTGATTATAA  CGCTAGTCGT  AGCACTGCTA  GCTAAATCAA
      251  AGGTTCCTGC  CATCCCAAC  GAACTTCAGA  AGATTATTTA  CAATCGCTAT
      25  301  CCTAAAGAAG  TCTTTTATTT  CGTGAAAACA  CACTCCCTGA  CTGTTAACGA
      351  ATTAATAATA  TTTATTAATT  GCTGAAAAG  CGGTACAGAC  CTGCCCTCGA
      401  ATTTACATAA  AAAAGCAGAG  GCTTTCGGGA  TCGATATTCT  AAAATCTATA
      451  GATTTAATCC  TGTTTCCAGA  GTTCGAAGAG  ATTCTTCTTC  AAAACTGCCC
      501  GTTATACTGG  CTCTCCCAT  TTATAGACAA  AACTGAATCT  GTTGCTGGGG
      30  551  AAATCGGATT  AAATAAAACA  CAAAAGTTT  ATGGTTTACT  TGGGCCCTTA
      601  GCGTTTCATA  AAGGATATAC  AACTATTTTC  CACTCTTATA  CACGCCCTCT
      651  ACTAACATTA  ATCTCAGAA  CACAGTATAA  GTTCTTATAT  AGTAAAGCGT
      701  CTAAGAATCA  ATGGGATTCT  CTTCTGTGA  AAAAAACCTG  CGAAGAAATA
      751  TTCAAGGAAC  TCCCCACAA  TATGATTTTC  CGGAAGGATG  TTCAAGGAAT
      35  801  CTCACAATTC  TTATTTCTTT  TCTTTTCTCA  TGGTATCACT  TGGGAACAGG
      851  CTCAGATGAT  TCAACTTATA  AATCCTGATA  ATTGGAAAT  GTTGTGTCAG
      901  TTTGATAAAG  CAGGAGGCCA  CTGTTCCATG  GCAACATTIG  GAGGCTTTT
      951  GAATACTGAA  ACAAAATATG  TCGATCCAGT  ATCCTCTAAC  TATGAACCTA
      1001  CAGTGAACCT  CATGACGTGG  AAAGAATTGA  AGGTTTACT  AGAGAAAGTA
      40  1051  AAAGAAAGTC  CTATGCACCC  AGCGAGTGCT  CTGTTCAGA  AGATATGCGT
      1101  AAATACAACG  CACCATCAA  ATCTGTTAAA  ACGATGGCAA  TTTGTTCTGA
      1151  ATACGAGTTC  ACAATGGACA  TCAAGCTTAC  CTCAGTATGC  TTTCCACGCC
      1201  CAAACCTACA  AACTAGAGAA  AAAAATAGAA  AGCAGTCTCC  CTATACGATC
      1251  TTCCCTATAA
```

- 45 The PSORT algorithm predicts inner membrane (0.7198).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 120A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 120B) and for FACS analysis.

- 50 These experiments show that cp6627 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

**Example 121**

The following *C.pneumoniae* protein (PID 4376629) was expressed <SEQ ID 241; cp6629>:

```

1 MSNITSPVIQ NNRSCNYYFE LKNSTTIHIV ISAILLCGAL IAFLCVAAPV
51 SYILSGALLG LGLLIALIGV ILGIKKITPM ISSKEQVFPQ ELVNRIRAHY
101 PKFVSDFVSE AKPNLKDILS FIDLLNQLHS EVGSSTNYNV SEELQOKIDT
151 FEGIARLKNE VRTASLKRLE SAASSRPLFP SLPKILQKVF PFFWLGEFIS
201 AGSKVVELHR VKKIGGSLEE DLSDYIKPEM LPTYWLIPLD FRPTNSSILN
251 LHITLVLARVL TRDVFQHLKY AALNGEWNLN HSDLNMTKQQ LFAKYHAAYQ
301 SYKHLSPQSL QEDEFYNLLL CIFKHRYSWK QMSLIKTVPA DLWENLCCLT
10 351 LDHTGRPDQM EFASLIGTLY TQGLIHKESE AFLSSLTLLS LDQFKTIRRO
401 STNIAMFLEN LATHNSTFRS LPPI TVHPLK RSVFSQPEED BSSLIG*
```

The cp6629 nucleotide sequence <SEQ ID 242> is:

```

1 ATGAGTAATA TAACCTCGCC AGTTATTCAA AATAATCGCT CTTGTAATTA
51 TTATTTTGAA TTAAAGAATT CAACCACTAT TCATATTGTT ATCAGTGCCA
15 101 TCTTACTCTG CGGAGCTTTG ATAGCTTTCT TGTGTGTAGC AGCTCCTGTT
151 TCCTATATTC TAAGTGGCGC ATTGTTAGGA TTAGGATTAT TAATAGCCTT
201 GATTGGTGTG ATTTTAGGAA TAAAAAAAT CACGCCTATG ATTTTCATCAA
251 AAGAACAAGT ATTCCTCCCA GAACTCGTAA ATAGAATCAG GGCGCACTAT
301 CCTAAATTTG TCTCTGATTT TGTTTCAGAA GCTAAACCAA ATCTTAAAGA
20 351 TCTCATAAGT TTTATTGATC TTCTAAATCA ATTGCACTCT GAAGTGGAT
401 CATCTACAAA TTACAACGTA TCTGAAGAAC TACAACAGAA AATAGATACG
451 TTCGAGGGTA TCGCACGCTT AAAAAATGAA GTCCGTACTG CTTCTCTTAA
501 AAGACTTGAA AGCGCTGCTT CTTCCCGTCC CCTCTTCCCC TCTTTACCAA
551 AAATCTTACA AAAGGTATTT CCAATTTTCT GGTTAGGAGA GTTTATTTCT
25 601 GCAGGCAGCA AGGTTGTAGA GCTCCATCGA GTTAAGAAAA TTGGAGGCAG
651 CCTCGAAGAA GACCTTAGTG ATTATATAAA ACCAGAGATG CTTCTTACCT
701 ATTGGTTGAT TCCTTTAGAT TTTAGACCAA CAAATTCCTC TATTCTAAAT
751 CTACACACAT TAGTTTAGC TAGAGTCTTA ACTCGTGATG TTTTCAACA
801 TCTTAAGTAT GCAGCATTA ATGGCGAGTG GAACCTGAAT CATAGTGATC
30 851 TAAATACTAT GAAACAGCAG CTCCTTGCTA AATATCATGC GGCGTATCAA
901 TCCTATAAAC ATCTATCTCA ACCCTCTCTT CAAGAGGATG AATTCTATAA
951 CCTGCTCTTG TGTATTTTGA AGCATAGGTA CTCGTGGAAG CAGATGTCCT
1001 TAATAAAAAC AGTCCCGGCT GATTTATGGG AAAACCTCTG TTGCTTGACT
35 1051 TTAGACCATA CAGGACGACC CCAAGACATG GAATTTGCC TCTTAATTGG
1101 TACTCTCTAC ACACAAGGCC TAATTCATAA AGAAAGCGAA GCATTTCTTT
1151 CTTCAATTGAC ACTCCTTAGT TTAGATCAGT TTAACACGAT CCGTCGTCAG
1201 TCAACCAATA TAGCGATGTT CCTTGAGAA TTAGCAACTC ATAATTCCAC
1251 CTTTAGAAGC TTACCACCTA TAACAGTCCA TCCACTCAAG AGAAGCGTCT
1301 TCTCCCAACC TGAAGAAGAC GAGTCTCCC TGCTGATAGG TTAG
```

40 The PSORT algorithm predicts inner membrane (0.5776).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 121A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 121B) and for FACS analysis.

These experiments show that cp6629 is a surface-exposed and immunoaccessible protein, and that it  
45 is a useful immunogen. These properties are not evident from the sequence alone.

**Example 122**

The following *C.pneumoniae* protein (PID 4376732) was expressed <SEQ ID 243; cp6732>:

```

1 MEMMSPFQQP EQCHFQVVGSL FLRPESLTRA RSDFEGRIV YEQMRVVEDA
51 AIRNLKKQT EAGLIFFFTDG EFRYSWDFD FMWGFHGVDR RRDSNDPEIG
50 101 VYLKDKISVS KHPFIEHFEP VKTFEKGNAK AKQTIPSPSQ FFHEMIFAPN
151 LKNTRKFYPT NQELIDDIVF YYRQVIQDLV AAGCRNLQLD DCAWCRLLDI
201 RAPSWSYGVDS HDRLQEIQLF FLWIHNLVLMK DRPEDLFVSL HVCRGDYQAE
251 FFSRRAYDSI EEPLFAKTDV DSYHYWALD DKYSGGAEPL AYVSGEKHVC
301 LGLISSNHSC IEDRDAVVSRI IYEAASYIPL ERLSLSPQCG FASCEGDHRM
```

-148-

351 TEEBQWKIA FVKEIAKEIW G\*

The cp6732 nucleotide sequence &lt;SEQ ID 244&gt; is:

```

1  ATGGAATGA TGAGCCCAT CCAACAACCT GAGCAATGTC ATTTTGATGT
5  51  TGTGGGAAGT TTCTTACGTC CTGAAAGTCT TACACGAGCA CGCTCTGATT
    101  TTGAAGAAGG AAGAATGTGC TATGAGCAGA TGCGAGTTGT CGAAGATGCT
    151  GCTATTCGTA ATCTCATAAA AAAGCAAACA GAAGCAGGTC TTATCTTTTT
    201  TACTGATGGG GAATCCGTA GGTATAGTTG GGATTTGAC TTTATGTGGG
    251  GATTCCATGG CGTGGATCGT CGCAGGGACT CTAATGACCC TGAAATTGGA
    301  GTGTATCTTA AAGATAAAAT CTCCGTATCA AAACATCCGT TTATAGAACA
10  351  TTTTCGATTT GTCAAAACTT TTGAGAAGGG AAATGCAAAA GCAAAACAAA
    401  CGATTCCTTC TCCATCACA TTTTCCATG AGATGATTTT TGCTCCTAAT
    451  CTGAAAAATA CTCGGAAGTT TTATCCTACG AATCAAGAGC TAATTGATGA
    501  TATTGTCTTT TATTATCGCC AAGTCATCCA AGATCTTTAT GCTGCAGGTT
    551  GTCGTAATTT GCAGTGGAC GATTGTGCTT GGTGTCGCCT CTTGGATATA
15  601  CGAGCGCCTT CTTGGTATGG TGTGATCT CATGACAGGT TGCAGGAAAT
    651  TTTAGAACAG TTTTATGGA TCCATAATTT AGTGATGAAG GATAGACCCG
    701  AGGATCTTTT TGTAACTCTG CATGCTCTGTC GTGGTGATTA TCAGGCCGAG
    751  TTTTCTCTA GACGAGCTTA TGATCTCTATA GAGGAGCCTT TATTTGCTAA
    801  GACCGATGTG GATAGTTATC ACTATTATTG GGCTCTTGAT GATAAGTATT
20  851  CAGGAGGTGC TGAGCCTTTA GCTTACGTCT CTGGAGAGAA ACACGCTCTGC
    901  TTGGGATTGA TCTCCAGCAA CCATCTTGT ATTGAAGATC GAGATGCTGT
    951  GGTTTCTCGT ATTTATGAAG CTGCGAGCTA CATTCCTTTA GAGAGACTTT
    1001  CTTTGAAGCC GCAATGTGGG TTTGCTTCTT GTGAGGGAGA CCATAGAATG
    1051  ACTGAAGAAG AACAGTGGAA GAAGATCGCC TTTGTGAAAG AGATTGCTAA
25  1101  AGAGATCTGG GGATAA

```

The PSORT algorithm predicts cytoplasm (0.2196).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 122A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 122B) and for FACS analysis.

- 30 These experiments show that cp6732 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 123

The following *C.pneumoniae* protein (PID 4376738) was expressed <SEQ ID 245; cp6738>:

```

35 1  VWLRFLLVLS YDEKEKDVVV VCNHSEPNIL GLPPEAVSQL IEELSDEGYS
    51  YLNVVRCDLS GETTVQQRLL LNADEGRSMT VVISELPEGH PDIRNLQLAS
    101  ERIFVSREKE AADAYASGCK VVAFDDEHLP WVSSHAYAE EIREKQEQTM
    151  QGSLTEEQLG ALLCNFVSTE KNLAFALDAV IKQSVWRFRN PDLFAYEREA
    201  LEASVTDALV SYVSNLDMIP YTSSQGIVIE DSSIVRTSQE HTLIVNCAAF
    251  DKLASQIEFL CPSDVLPISG KDPLISDDRD EELNPKVSSA ADSKDKT*

```

40 The cp6738 nucleotide sequence &lt;SEQ ID 246&gt; is:

```

1  GTGTGGCTGC GCTTTTACT TTTAGTGTC TATGATGAGA AGGAGAAAGA
51  CGTAGTTGTC GTTTGTAATC ATTCTGAACC TAATATCCTC GGCCTGCCTC
101  CTGAAGCAGT CTCTCAGCTT ATTGAAGAGC TTAGCGATGA AGGCTATAGC
45  151  TATCTGAATG TAGTCCGTTG TGATCTCTCC GGGGAGACTA CGGTTCAACA
    201  ACGTCTGCTA TTGAATGCCG ATGAAGGGAG ATCTATGACG GTGGTGATCT
    251  CAGAGCTTCC TGAAGGGCAC CCCGATATTC GGAATTTGCA GTTGGCATCC
    301  GAAAGAATTT TTGTTTCTCG TGAAAAGAGAA GCTGCTGATG CCTATGCTTC
    351  AGGATGTAAA GTGGTCGCTT TCGATGATGA GCATCTCCCT TGGGTCTCCA
    401  GTCATATTGC CTACGCGGAG GAGATCAGAG AGAAACAAGA ACAAACAATG
50  451  CAAGGGTCTT TAACGGAAGA GCAGTTAGGA GCACTCCTCT GCAACACAGT
    501  CTCCACAGAG AAAAATCTAG CCTTTGCTCT AGACGCCGTG ATAAAACAGT
    551  CTGTGTGGAG ATTCCGCAAT CCGGATCTTT TTGCTTATGA GAGAGAAGCT
    601  CTAGAGGCTT CAGTAACAGA TGCTTTAGTA TCTTACGTTT CAAATTAGA
    651  CATGATACCG TACACAAGTT CTCAGGGCAT AGTCATAGAA GATAGTAGTA
55  701  TCGTCCGTAC CTCTCAAGAG CATACTCA TTGTGAAGT TGCAGATTTC

```

-149-

```

751 GATAAGTTAG CGAGCCAAAT AGAGTTCTTA TGCCCCAGTG ACGTGTGCGC
801 CATTTCTGGT AAAGACCCTT TGATTCTCTGA TGATGAGGAT GAGGAAGTGA
851 ATCCTAAAGT TTCATCTGCT GCAGACTCTA AAGATAAAAC CTAG

```

The PSORT algorithm predicts cytoplasm (0.1587).

- 5 The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 123A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 123B) and for FACS analysis.

These experiments show that cp6738 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

## 10 Example 124

The following *C.pneumoniae* protein (PID 4376739) was expressed <SEQ ID 247; cp6739>:

```

1 MTHCLHGWFV VVRHHFVQAF NFSRPLYSRI THFALGVIKA IPIVGHVLMG
51 VDWLLSHCFE RGVSHPGFPS DIAPILKVEK IAGRDHISRI ENQLKSLRKT
101 IEVEDLDKVH GQYQENPYAD MASSEVLKLD KGVHVSELGK AFSRVRNRIT
15 151 RSYSYAPTPO LDSIAIVGID LVSPBQENL VRLANEVIQL YPKSKTTLYL
201 LIDFNKEWVG DISSDKKKQL RSLGLHSEVQ CLSVLEPQGA EGEDTKHFDL
251 MVGCGYKDSY LREGKILQQA LGTSLGTVPW VNMHTLPSR YRSRLSLPIN
301 TEKDKTELYK EISRTHQLH TLGMGLGAQD SGLLLDRQRL HAPLSQGS HC
351 HSYLADLTBE ELKILLFSAF VDAKNISKKE LREVSILNFAN DTSVECGCAF
20 401 YF*

```

The cp6739 nucleotide sequence <SEQ ID 248> is:

```

1 ATGACTCATT GCTTACATGG TTGGTTTCTT GTAGTTCGTC ATCATTCTGT
51 GCAGGCGTTT AATTCTCTAC GTCCTTTATA TTCTCGAATT ACCCACTTCG
25 101 CTTTAGGGGT GATTAGGCC ATCCCCATTG TAGGGCATCT TGTTATGGGA
151 GTCGATTGGT TGATCTCTCA TTGCTTCGAG AGGGGAGTCT CACACCCTGG
201 GTTCCCTTCA GATATGCTC CTATACTGAA AGTAGAAAAG ATCGCGGGCC
251 GAGATCATAT TTCTAGAATC GAAAATCAGC TAAAGAGCCT TAGGAAAAC T
301 ATCGAGGTG AAGATCTAGA TAAAGTCCAC GGGCAATATC AAGAGAATCC
351 TTATGCAGAT ATGGCCTCTA GTGAGGTCTT TAAACTCGAT AAGGGAGTTC
30 401 ATGTTAGCGA GCTTGGCAAA GCCTTTTCTA GAGTTCGCAA TCGCATCACC
451 AGATCCTATA GTTATGCCCC TACTCCTCAG TTGGACTCTA TAGCTATTGT
501 TGGTATAGAT CTCGTCAGTC CTGAAGAACA AGAGAATTTA GTACGCTTGG
551 CGAATGAGGT CATTCAACTC TATCCCAAAT CAAAGACAAC TCTATATCTT
601 CTTATCGATT TTAATAAGGA GTGGGTAGGG GATATCTCCT CTGATAAGGA
35 651 AAAACAGCTC CGTTCTCTAG GTCTACATTC TGAAGTTCAG TGTCPTTCCG
701 TCTTGGAAAC TCAGGGTGCC GAGGGCGAAG ATACGAAACA CTTTGACCTT
751 ATGGTTCGGT GTTATGGGAA GGATTCTTAC TTAAGGGAGG GTAAAAATTT
801 ACAGCAGGCC CTAGGGAATT CGTTAGGTAC TGTTCCTCTG GTGAATGTTA
851 TGCACACATT GCCATCTAGG TATAGATCTC GGCTTTCTCT ACCTATAAAT
40 901 ACCGAAAAGG ATAAGACAGA GCTTTATAAA GAGATTCTCT GTACACACCA
951 TCAGTTGCAT ACTTTGGGAA TGGGACTTGG AGCCCAGGAT TCAGGATTGC
1001 TCTTAGACCG GCAACGACTC CATGCTCCTT TATCTCAAGG GTCTCACTGC
1051 CATTCCTATC TTGCAGATCT CACCCATGAA GAGCTGAAA TTTTGTATT
1101 TTCAGCATT GTGGATGCTA AGAACATAAG TAAGAAAGAG CTTCTGTGAGG
45 1151 TATCTCTAAA TTTTGCTAAC GATACTTCCG TAGAGTGTGG CTGCGCTTTT
1201 TACTTTTAG

```

The PSORT algorithm predicts inner membrane (0.2190).

- 50 The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 124A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 124B) and for FACS analysis.

-150-

These experiments show that cp6739 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 125

The following *C.pneumoniae* protein (PID 4376741) was expressed <SEQ ID 249; cp6741>:

```

5      1  MASCLSAWFS  IVREHFYRAF  DFSLPFCARI  TEFVLGVIKG  IPVVGHIIVG
      51  IEWLVSRYLE  SFVTKPTFVS  DVVSLKTEK  VAGRDHIARV  VETLKRQ RVA
     101  VAPEDEDKVH  GKIPVHPFGG  IQPVEVLTL  Y  PEVQDATLGL  AFSKIRNRVR
     151  QAYLQAPRPK  LQKIYIIGND  MNPFEVDDFL  HLA RL CNETQ  RLYPDATISL
     201  YLTASGGRNA  MDKKNRKL  LS  DCELNPKIAC  LDFNQGDVVK  QATCDCWMVY
    10  251  HGENDQGT  LN  QIQEELKSG  EETPWTHVGQ  KPLSQSLWDF  SPFSSLEMKG
     301  DKEKALEYSE  LEKEQLYSRL  VYVGERSSVL  SLGFGDSRSG  ILMDPKRVHA
     351  PLSEGHYCHS  YLADLENPGL  QKTILAAFLN  PKELSSTILQ  PISLNLILNS
     401  KTYLRQHFGF  FERMSRSDRN  VVVVVCD  SWW  GTDWKEEPSF  QHFIMELECR
     451  GYSHFNIFAF  RSNSMCVEER  RILNESSQEK  AFTMIFCEDS  VSQGDIRCLH
    15  501  LASEGMLCGK  ECVAVDVYTS  GCANFMMEEV  LTLERESNLW  NRKHGLWKRE
     551  VRKQRQEAAL  DQDSEIYVC  NQLTAQQNFA  CS*

```

The cp6741 nucleotide sequence <SEQ ID 250> is:

```

      1  ATGGCTTCTT  GTTTATCTGC  CTGGT TTTCT  ATAGTTCGTG  AGCACTTTTA
     51  TCGAGCCTTT  GATTTTCTCT  TGCCGTTT  TG  TGCTCGTATT  ACGGAATTTG
    20 101  TATTAGGGGT  CATCAAGGGG  ATCCCTGT  TG  TGGGTCACAT  TATTGTTGGG
     151  ATAGAGTGGC  TCGTTTCTAG  GTATTTAGAG  AGTTTCGTGA  CCAAGCCGAC
     201  ATTTCTCTCT  GATGTGGTGA  GTCTTCTGAA  AACAGAGAAA  GTTGCTGGTC
     251  GCGATCACAT  TGCTCGTGTA  GTGGAGACTT  TGAAGAGGCA  GAGAGTCGCT
    301  TGGGCTCCTG  AAGATGAGGA  TAAGGTCCAT  GGGAAAGATT  CTGTGCATCC
    25 351  TTTCGGGGGA  ATCCAACCTG  TAGAAGTTCT  CACTCTCTAT  CCCGAAGTTC
     401  AAGATGCAAC  GTTAGGGCTT  GCCTTCTCTA  AAATTCGTAA  TCGTGTAAGA
     451  CAGGCGTATT  TGCAAGCTCC  ACGGCCAAA  CTGCAGAAGA  TTTACATCAT
     501  AGGAAACGAT  ATGAATCCTT  TTGAAGTTGA  CGACTTCTTG  CATCTAGCCC
     551  GTCTCTGTAA  TGAAACTCAA  AGACTCTATC  CTGACGCTAC  GATTTCTCTA
    30 601  TATCTAACAG  CTTCTGGTGG  TCGCAATGCT  ATGGCAAAA  AGAATCGGAA
     651  GTTACTTAGT  GATTGCGAAC  TAAACCCCAA  GATTGCTTGT  TTGACTTTA
     701  ATCAGGGTGA  TGTAGTCAAA  CAAGCAACTT  GTGACTGTTG  GATGGTGTAT
     751  CATGGGGAGA  ATGATCAAGG  TACGTTGAAT  CAGATTCAGG  AAGAGTTAGA
     801  AAAGTCAGGG  GAGGAAACCC  CTTGGATTCA  TGTGGGGCAA  AAGCCTCTTT
    35 851  CACAATCCTT  GTGGGATTTT  TCTCCATTTT  CATCTTTGGA  GATGAAGGGA
     901  GATAAAGAGA  AAGCTCTAGA  GACTCTGAA  TTAGAAAAAG  AACAGCTATA
     951  TTCTCGATTG  GTATACGTAG  GAGAGCGCTC  TTCCGTTCTT  AGTTTGGGGT
    1001  TTGGAGATAG  TCGGTCAGGG  ATCTTGATGG  ACCCAAAACG  GGTGCATGCT
    1051  CCCTTATCTG  AAGGGCATT  A  TTGTCAATCC  TACCTTCAG  ACTTAGAAAA
    40 1101  TCCCGGGTTA  CAAAAACAA  TTTAGCGGC  ATTTCTGAAT  CCTAAGGAGT
     1151  TGAGCAGTAC  CATACTGCAA  CCTATATCTC  TAAATCTTAT  CTTAAATAGC
     1201  AAAACTTACT  TAAGGCAGCA  CTTTGGCTTT  TTTGAGAGGA  TGAGCAGAAG
     1251  TGATCGCAAT  GTGGTTGT  CG  TTGTATGTGA  TTCTTGGTGG  GGTACCGACT
     1301  GGAAGGAGGA  GCCAAGCTTC  CAACACTTTA  TTATGGAGCT  AGAGTGTCTG
    45 1351  GGGTATTCGC  ACTTCAATAT  TTTTGCCTTT  AGATCTAATA  GCATGTGTGT
     1401  AGAAGAACGT  AGGATCTTAA  ATGAAAGTTC  TCAAGAGAAA  GCCTTTACCA
     1451  TGATTTTCTG  TGAGGATTCA  GTATCTCAAG  GAGATATCCG  CTGTTTGCAT
     1501  TTGGCGTCTG  AAGGAATGCT  TTGTGGTAAA  GAGTGCTATG  CTGTTCGATG
    50 1551  CTATACGTCA  GGATGCGCGA  ACTTTATGAT  GGAAGAAGTC  TTAAC TTTGG
     1601  AGCGAGAATC  TAATCTGTGG  AATAGAAAAG  ATGGTCTTTG  GAAAAGAGAA
     1651  GTTAGAAAAC  AGAAACAAGA  AGCTGCTTTG  GATCAAGACG  AGAGCGAGAT
     1701  TTACGTTTGT  AATCAGCTGA  CGGCGCAACA  GAACTTCGCT  TGTTC TTGA

```

The PSORT algorithm predicts inner membrane (0.2869).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 125A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 125B) and for FACS analysis.

These experiments show that cp6741 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 126

The following *C.pneumoniae* protein (PID 4376742) was expressed <SEQ ID 251; cp6742>:

```

5      1  LFVSNFIFFV  VMPIPYISSW  ISTVRQHFKV  AFDFSRPFCS  RVTNFGALGVI
      51  KAIPVGHIV  MGMEWLVSSE  VAGIITRSSF  TSDVVQIVKT  EKALGRDHIS
     101  RVAEILQRE  GTITPENQDK  VHGFPPVCP  GRLKSEETLK  LKPGEREGTL
     151  DTVFSPITR  VTRAYLQAPR  PEIRTISIV  SKLKTPODFS  QFVSLANETQ
     201  RLHPEALVCL  YLTGLNRESQ  MCDTTTAEKK  QYLHNSGLDS  RIQCKDSKED
10    251  DAGSPENPEL  WIGYYSREQQ  HNIDGQYIQ  CLGKSADPIP  WIHVTEDTKD
     301  FYYPPNFTSY  SHTRQSTDP  SPPRLPESEG  DKDSLYGQLS  RSYHHEYMLG
     351  LGLKPEDAGL  LMDPDRIYAP  LSQGHYCHSY  LADIENEDLR  TLVLSPLDLP
     401  GNLSEEDLRP  VAFNIARLPL  ELDSLFFRLV  AGQQEGRNIV  TLAHGTPRPE
     451  DLDPDSMNIL  TRRLQMSGYS  YLNIFYSKSR  KMTVKERQFF  GDRSEKGSFT
15    501  LILFEDPISA  ADFRCLQLAA  EGMVAKDLPS  VADICASGCS  CIQFSEMOSP
     551  QAIEYRQWEA  RVEDEAGER  REPVIYSQDQ  LSSMLTTQQN  FVFSLDAVVK
     601  QAIWRFRSKG  LLTMRKALG  EEFLTAIFY  LGSQERNENM  GKRTTEEHEV
     651  VISFEELDRM  VQVLPAEVPA  DSGNDPTRPV  PNPDSNPDS  QNEGS*

```

The cp6742 nucleotide sequence <SEQ ID 252> is:

```

20      1  TTGTTTGT  CTAATTTAT  TTTTFTTGT  GTTATGCCAA  TTCCCTATAT
      51  TTCTTCTT  ATTTCACCG  TTCGACAGCA  TTTTGTAAAG  GCGTTTGATT
     101  TCTCTCGT  CTTTCTTCT  AGGGTTACGA  ATTTTGCTTT  AGGGGTCATC
     151  AAGGCCAT  CTATTGTAG  ACATATTGTC  ATGGGGATGG  AGTGGTGTAG
     201  TTCTTCCT  GTTGCCGGA  TTATTACTAG  GTCCTCCTTT  ACCTCAGATG
25    251  TCGTTCAG  TGTAAAGAC  GAGAAGGCGT  TAGGTCGAGA  TCATATATCT
     301  CGAGTGGC  AGATATTGCA  AAGAGAAAGG  GGGACCATAA  CTCCTGAGAA
     351  TCAAGATA  GTGCATGGA  AGTTTCTGT  CTGTCCTTTT  GGTCTGTTAA
     401  AATCCGAG  AACTTTAAA  CTTAAGCCGG  GAGAAAGAGA  GGGAACTTTA
     451  GATACTGT  TTTCTCCGAT  TCGCACGCGC  GTGACTCGTG  CGTACTTACA
30    501  GGCCCCCG  CCCGAAATAC  GTACGATTTC  TATTGTGGGT  TCGAACTPTA
     551  AAACCTCT  AGATTTCTCG  CAATTTGTGA  GTCTCGCGAA  TGAAACGCAG
     601  AGACTGCA  CTGAAGCGTT  AGTTTGTCTG  TATTGTACAG  GCTTGAATCG
     651  CGAATCTC  ATGTGCGATA  CAACTACTGC  AGAGAAGAAG  CAGTACCTAC
     701  ATAACTCAG  TCTCGACTCT  AGAATCCAGT  GCAAAGACAG  TAAAGAAGAC
35    751  GACGCTGG  CTCCTGAAA  TCCCGAACTT  TGGATTGGCT  ATTATTCACG
     801  AGAGCAAC  CATAATATAG  ACGGGCAGTA  TATTTCAGCAG  TGTCTAGGGA
     851  AGAGTCAGA  TCCAATTCCT  TGGATTTCATG  TTAAGTGAAGA  CACAAAGGAT
     901  TTTTATTA  CACCAAACTT  TACTTCATAC  TCACATACAA  GACAATCTAC
     951  AGACCCAA  TCGCCACCAA  GACTCCCTGA  AAGTGAGGGG  GATAAGGATT
40   1001  CCTTGTAC  ACAACTGAGT  CGATCGTATC  ACCATGAGTA  TATGCTTGGT
     1051  TTGGGATT  AACCAGAGGA  TGCAGGACTC  CTGATGGACC  CGGATAGAAT
     1101  CTATGCTC  CTATCCCAAG  GGCATTATTG  TCATTCTTAC  CTTGCGGATA
     1151  TAGAAAAT  GGATCTACGA  ACTTTAGTCC  TTTTCGCTTT  CCTAGATCCT
     1201  GGCAATCT  GTAGCGAGGA  TCTTCGTCTT  GTAGCATTCA  ATATCGCTAG
45   1251  ATTGCCAT  GAATTGGACT  CGTTATTTT  CCGCCTTGTT  GCGGGTCAGC
     1301  AAGAAGGG  AAACATAGTT  ACCCTTGCCC  ACGGAACTCC  TCGTCCAGAA
     1351  GATCTTGA  CTGACTCAAT  GAACATCTTG  ACCAGAAGAT  TACAAATGTC
     1401  TGGATATA  TATTTGAACA  TTTTCTCCTA  TAAATCACGG  AAAATGATTG
     1451  TAAAAGA  TCAGTTCTTT  GGAGATCGTT  CTGAAGGGAA  GTCTTTCACA
50   1501  TTGATCTT  TTGAGGATCC  CATTAGTGCA  GCAGATTTCC  GTTGTTTGCA
     1551  GCTAGCTG  GAAGGTATGG  TTGCTAAGGA  TCTCCCAGC  GTAGCAGATA
     1601  TTTGTGCT  TGGATGTTCC  TGCATTCAGT  TTTCTGAGAT  GCAGAGTCCT
     1651  CAGGCTATT  AATATAGACA  ATGGGAGGCA  CGTGTCAAG  ATGAAGCAGG
     1701  AGAAGAAG  AGAGAAACCAG  TAATTTATTC  TCAGGATCAA  TTGAGCAGCA
55   1751  TGCTCACT  ACAACAGAA  TTTGTATTTT  CTCTAGATGC  TGTGGTAAAA
     1801  CAGGCGAT  GGAGATTCCG  TTCGAAAGGT  CTTCTTACTA  TGGAAAGAAA
     1851  GGCAC TAGG  GAGGAGTTCT  TAACTGCGAT  ATTTTCTTAT  TTAGGGAGTC
     1901  AGGAGCGT  TGGAATATG  GGGAAAAGAA  CTACCGAAGA  ACATGAGGTC
     1951  GTTATCAG  TCGAAGAGCT  AGATCGCATG  GTGCAAGTCC  TCCCAGCCGA
60   2001  AGTCCCTG  GATTCAAGCA  ATGATCCTAC  GCGTCCCGTT  CCTAATCCAG
     2051  ATAGTAACC  TGATTCTCTG  CAAAATGAAG  GCAGTTAG

```

The PSORT algorithm predicts inner membrane (0.2338).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 126A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 126B) and for FACS analysis.

- 5 These experiments show that cp6742 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

#### Example 127

The following *C.pneumoniae* protein (PID 4376744) was expressed <SEQ ID 253; cp6744>:

```

10      1  VIQHLNLFAL  EETPSISVQY  QEQEKLSPCD  HSPBIGKKKR  WNKLESFSTY
      51  CSLFMSVKDH  YKLNLIQONS  LSGWLLDPYR  VCAPLSSPYS  CPSYLLDLQN
     101  KELRRSLLST  FLDPKNLTSE  TFRSVSINFG  NSSFGQRWSE  FLSRVLHDEK
     151  EKHVAVVCND  AKLLEEGLSP  EALSILLEEDL  RESGYSYLN  LSVSPEGVSK
     201  VQERQILRRD  LQGRSFTVMI  TDLPLGSEDI  RSLQLASDRI  LVSSSLDAAD
     251  ACASGCKVLV  YENPNASWAQ  ELENFYKQVE  RRR*

```

- 15 The cp6744 nucleotide sequence <SEQ ID 254> is:

```

      1  GTGATACAAC  ATCTTCTAAA  CTTTGCTCTA  GAAGAGACCC  CTTCCATTTC
     51  CGTGCAATAC  CAAGAACAAG  AGAAGCTCTC  TCCGTGCGAT  CATTCCCCAG
    101  AAATAGGTAA  AAAGAAAAGA  TGAATAAGC  TGGAAATCCTT  CTCCACGTAT
    151  TGTTCCTCTGT  TTAGTCTCTGT  TAAGGATCAT  TATAAGCTGA  ATCTAGGAAT
    201  TCAGAATTCC  CTGTCAGGGT  GGCTTCTGGA  TCCCTATAGG  GTTTGCGCGC
    251  CTTTATCTTC  ACCGTACTCG  TGTCTTCTCT  ATCTTTTAGA  TTTGCAAAAC
    301  AAAGAGCTAC  GTCGTTCCCT  TCTGTCAACG  TTTCTAGACC  CTA AAAATCT
    351  CACTAGCGAA  ACATTCGGTT  CTGTCTCTAT  AAAC TTTGGC  AACTCTTCGT
    401  TTGGACAGAG  ATGGTCAGAG  TTTCTATCTC  GTGTTCTGCA  CGACGAGAAA
    451  GAAAAGCAGC  TAGCTGTTGT  TTGTAATGAT  GCAAAACTTC  TGGAAGAAGG
    501  ATTGTCCCCA  GAGGCATTGT  CTCTATTAGA  AGAAGACTTA  AGAGAATCAG
    551  GGTATTCGTA  TCTAAACATT  CTCTCGGTGA  GCCCCGAAGG  AGTCTCCAAG
    601  GTTCAGGAAC  GTCAGATTCT  AAGGCGAGAT  CTCCAAGGAC  GGTCTTTTAC
    651  TGTTCATGATT  ACAGATCTTC  CTTTAGGTAG  CGAAGATATC  CGTAGTTTAC
    701  AATTAGCCTC  GGATAGGATT  TTAGTCTCCA  GTTCTCTTGA  TGCCGCGGAT
    751  GCATGTGCTT  CGGGATGTAA  AGTCTTAGTC  TACGAAAATC  CAAATGCATC
    801  CTGGGCTCAG  GAATTGGAGA  ACTTCTACAA  ACAAGTTGAG  AGAAGAAGGT
    851  AG

```

The PSORT algorithm predicts cytoplasm (0.3833).

- 35 The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 127A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 127B) and for FACS analysis.

These experiments show that cp6744 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

#### 40 Example 128

The following *C.pneumoniae* protein (PID 4376745) was expressed <SEQ ID 255; cp6745>:

```

      1  VACPSSSWF  TVVRQHFVNA  FDFTHPVCSR  ITNFALGIK  AIPVLGHIVM
     51  GIEWLISWIP  RHTVRHGMFT  SDVSSAIKVE  QTRGHNCLAP  LEAYLSSLRV
    101  PISQEDLGKV  HGRTPEDPFV  DITPTEIVQL  LPDEELSTVD  EALQGVRSRL
    151  TYAYRSVEKP  MIQDLALVGF  GLRDSADLIN  FVRLANGVQN  HYPHTKVKLY
    201  LAKNLADVWD  CEISEEEKGQ  LRALGLDPKI  ESISLTSAGL  PSVPEVATVD
    251  FMITCYGKDQ  EVQDP*

```

The cp6745 nucleotide sequence <SEQ ID 256> is:

```

1  GTGGCTTGTC CAAGTATTTC TTCTTGGTTT ACTGTCGTTT GACAGCATT
51  TGTAACGCC TTGATTTTCA CCCATCCCGT TTGTTCTCGG ATTACAAAT
101  TTGCTTTGGG GATCATTAAG GCAATCCCG TATTAGGACA CATTGTCATG
151  GGAATCGAGT GGTGATTTTC CTGGATTCCC AGACACACCG TTCGTCATGG
201  AATGTTTACT TCTGATGCTT CTAGTGCTAT TAAAGTAGAA CAAACACGGG
251  GTCATAATTG TTTAGCTCCC CTAGAAGCCT ATTTAAGTAG CTTGAGAGTC
301  CCCATTTCCT AAGAAGATCT AGGCAAAGTA CACGGGAGAA CCCCAGAAGA
351  TCCCTTCGTA GATATCACAC CCACAGAAAT TGTCCAACCT CTCCTGATG
401  AAGAACTCTC TACTGTAGAT GAGGCACTGC AAGGCGTTTC TAGTAGGTTA
451  ACCTATGCCT ATAGGTCCGT AGAGAAACCT ATGATTCAGG ATCTTGCTCT
501  TGTGGGTTTT GGTCTCCGAG ATTCTGCGGA CCTCATAAAT TTCGTGCGTC
551  TTGCTAATGG CGTGCAGAA CACTATCCCC ATACTAAAGT GAAGCTCTAT
601  TTAGCGAAGA ACTTGCAGTA TGTCTGGGAC TGTGAAATTT CTGAAGAGGA
651  AAAAGGGCAA CTCCGAGCTC TAGGTTTAGA CCCTAAAATA GAGAGTATAT
701  CCCTTACGAG TGCAGTCTT CCTTCAGTGC CAGAAGTCGC TACTGTCGAT
751  TTTATGATTA CCTGTACGG GAAAGATCAG GAAGTCCAAG ATCCCTAG

```

The PSORT algorithm predicts inner membrane (0.2253).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 128A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 128B) and for FACS analysis.

These experiments show that cp6745 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

#### Example 129

The following *C.pneumoniae* protein (PID 4376747) was expressed <SEQ ID 257; cp6747>:

```

1  MMKQGVGQDA KELYTFLSRG NEHYQPCLWF SLKEELGFLF DEKMLCAPLS
51  EDHYCHSYLV DLVDQHLKDL ILSMFLDPQN ISAGELLKVS INVGDSPSPL
101  QQKDFLSMVL RDETGNVNVV VFKGVLSPAL TQVCKLVEEL NSKDYSYLN
151  FSCHGDSPPQ LLFRKELEGT SGRTFTVICA LYLGDFTMRS LQLASERIMV
201  SREFDLVDAY AARCKLLKID HTNWRPGTFS RHADFADAVID VSAGFNSREF
251  KLITQANQGI LESGELPLPS KTFWEGFLAF CDRVTVTRHF IPMLDAATKQ
301  AVWTHKHPSL IDKECEALDL KTQCLPSIVS YLEYVTNSHE RTSKGPFIQK
351  EIIADCSPLK EALFPGSDED VPSTSEDPDS DHPSDLBDS*

```

The cp6747 nucleotide sequence <SEQ ID 258> is:

```

35  1  ATGATGAAAC AAGGAGTCGG GCAGGATGCT AAAGAGCTAT ACACATTTCT
51  ATCTCGTGGG AATGAGCATT ACCAACCCTG TCTATGTTTC AGTCTCGAAG
101  AGGAATCGCG ATTCTTTTTC GATGAAAAA TGCTCTGCGC CCCTCTATCT
151  GAGGATCACT ATTGCCACTC GTATCTTGTA GATCTAGTGG ATCAACATTT
201  AAAGGATTTA ATATTATCGA TGTTTTTAGA TCCTCAGAAT ATCTCAGCAG
40  251  GAGAACTCCT CAAGGTCTCT ATAAACGTTG GAGATTCTTT TTCTCCTCTA
301  CAACAGAAAG ATTTCTCTCT GATGGTCTTA CGTGATGAAA CGGGAAAAAA
351  CGTCGTCGTG GTTTTTAAAG GAGTCTCTCT CTTACCCGCA ACCCAAGTCT
401  GCAAATTAGT AGAGGAATTG AACTCTAAGG ACTACTCCTA CCTCAATATA
45  451  TTTTCTTGTC ACGGAGATAG TAGTCCTCAG CTTTTATTTC GTAAGGAATT
501  AGAGGGAACT TCAGGCGGTT ATTTTACAGT GATTTGCGCT TTATATCTAG
551  GGGATACAGA CATGCGTAGT TTACAACCTG CTTCTGAAAG GATCATGGTC
601  TCTAGAGAGT TTGATCTTGT AGATGCCTAT GCTGCAAGAT GCAAGCTCTT
651  GAAAATCGAT CATACAAATT GGAGACCTGG AACTTTCAGT CGCCACGCCG
701  ATTTTCGAGA TGCTGTAGAC GTATCAGCAG GATTTAACTC AAGAGAATTT
50  751  AAACATGATTA CGCAGCGGAA TCAAGGGATC CTAGAGTCTG GAGAACTCCC
801  GCTCCCTTCA AAAACCTTCT GGGAAGGATT CTTAGCATTC TGTGATCGAG
851  TGACTGTACAC GAGACACTTC ATTCCAATGT TAGACGCCGC TATAAAGCAA
901  GCGGTATGGA CTCATAAACA TCCCAGCTTG ATAGATAAAG AGTGTGAAGC
951  CCTAGACTTG AAAACACAGT GCTTGCCATC TATCGTATCG TACCTTGAAT
55  1001  ATGTCACAAA CTCTCACGAA AAAACATCGA AAGGCCCGTT CATACAAAAA
1051  GAGATTATCG CAGACTGTTC TCCTCTTAAA GAGGCGCTCT TCCCAGGTTT

```



-154-

1101 TGATGAAGAT GTTCCCTCTA CCTCTGAGGA TCCTTCAGAT GATCATCCTT  
1151 CGGATCTTGA AGACTCTTAA

The PSORT algorithm predicts inner membrane (0.1447).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 129A) and also as a his-tagged product. The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (Figure 129B) and for FACS analysis.

These experiments show that cp6747 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 130

The following *C.pneumoniae* protein (PID 4376756) was expressed <SEQ ID 259; cp6756>:

1 MASGIGGSSG LGKIPPKDNG DRSRSPSPKG ELGSHEISLP PQEHGEEGAS  
51 GSSHIHSSSS FLPEDQESQS SSSAASSPGF FSRVRSGVDR ALKSFGNFFS  
101 ABSTSQARET RQAFVRLSKT ITADERRDVD SSSAAATEAR VAEDASVSGE  
151 NPSQGVPEFS SGPEPQRLFS LPSVKKQSG LRLVQTVRDR IVLPSGAPPT  
201 DSEPLSLYEL NLRSLSSLRQE LSDIQSNDQL TPEEKAEATV TIQQLIQITE  
251 FQCGYMEATQ SSVSLAEARF KGVETSDEIN SLCSRLTDPE LQELMSDGDS  
301 LQNLLEDFTAD DLEAALSHTR LSFSLDDNPT PIDNNPTLIS QKEPIYEEIG  
351 GAADPQRTRE NWSTRLWNQI REALVSLGGM ILSILGSILH RLRIARHAAA  
401 EAVGRCCTCR GEECTSSSEED SMSVGSPEI DETERTGSPH DVPRRNGSPR  
451 EDSPLMNALV GWAHKGAKT KESSESSTPE ISISAPIVRG WSQDSSVSFI  
501 VMEDDHIFYD VPRRKDGIYD VPSSPRWSPA RELEEDVFGD YEVPTSAEP  
551 SKDKNIYMTF RLATPAIYDL PSRPGSSGSS RSPSSDRVRS SSPNRRGVPL  
601 PPVPSAMSE EGSIEDMSG ASGAGESDYE DMSRSPSPRG DLDEPIYANT  
651 PEDNPFTQRN IDRILQERSG GASAPFVEPI YDEIPWIHGR PPATLPRPEN  
701 TLTNVSLRVS PGFGPEVRAA LLSSEVSAMV VEASIVPPT EPGDGESEYL  
751 EPLGLVATT KILLQKGWPR GESNA\*

The cp6756 nucleotide sequence <SEQ ID 260> is:

1 ATGGCATCAG GAATCGGAGG ATCTAGTGGA TTAGGAAAGA TTCCACCTAA  
51 AGATAATGGG GATAGAAGTC GATCGCCCTC TCCTAAGGGA GAACCTGGCA  
101 GCCACGAGAT TTCCCTGCCT CCTCAAGAAC ATGGAGAGGA AGGAGCTTCA  
151 GGATCTTCGC ATATACATAG CAGTTCTCTT TTTCTACCAG AAGATCAGGA  
201 GTCTCAGAGC TCTTCTTCGG CAGCTTCTAG CCCGGGATTT TTTTCTCGCG  
251 TACGTTCTGG GGTAGACAGG GCCTTAAAT CATTTGGCAA CTTTCTTTC  
301 GCAGAGTCTA CGAGTCAAGC GCGTGAAACG CGACAAGCTT TTGTTAGATT  
351 ATCAAAAACC ATCACCAGCG ATGAGAGACG GGATGTCGAT TCATCAAGTG  
401 CTGCTGCTAC AGAAGCCCGA GTGGCAGAGG ACGCGAGTGT TTCAGGCGAA  
451 AATCCTTCTC AGGGGGTTCC AGAAACCTCT TCTGGACCAG AACCTCAGCG  
501 TTTATTTTCT CTTCCTTCAG TAAAAACA GAGCGGTTTG GGTGCGTTGG  
551 TACAGACAGT TCGCGATCGC ATAGTACTTC CTAGTGGGGC TCCACCTACA  
601 GACAGCGAGC CTTTAAGTCT CTACGAGCTA AACCTCCGTT TGAGTAGTTT  
651 ACGTCAGGAG CTCTCTGACA TACAAAGTAA TGATCAGTTG ACTCCAGAGG  
701 AAAAAGCAGA AGCCACAGTT ACCATACAAC AGCTGATCCA AATTACAGAA  
751 TTCCAATGCG GCTATATGGA GGCAACACAA TCTTCGGTAT CTCTAGCAGA  
801 AGCTCGTTT AAGGGGGTAG AAACCTAGTGA TGAGATCAAT TCCCTCTGTT  
851 CAGAACTGAC AGATCCTGAG CTTCAAGAAC TCATGAGTGA TGGAGACTCT  
901 CTTCAAAACC TATTAGATGA GACTGCCGAC GATTTAGAAG CTGCTTTGTC  
951 CCATACTCGA TTGAGTTTCT CTTTAGACGA TAATCCAAC CCGATAGACA  
1001 ATAAATCCAAC TCTGATTCT CAAGAAGAGC CTATTATGA GGAAATCGGA  
1051 GGAGCTGCAG ATCTCAAG AACTCGGGAA AACTGGTCTA CAAGATTATG  
1101 GAATCAGATT CGCGAGGCTC TGGTTTCTCT TTTAGGAATG ATTTTAAGCA  
1151 TTCTAGGGTC CATCTGCAC AGGTTCGCTA TTGCTCGTCA TGCAGCTGCT  
1201 GAAGCAGTGG GTCGTTGTTG CACGTGCCGA GGAGAAGAGT GTACTTCTTC  
1251 TGAAGAGGAC TCGATGTCGG TGGGGTCTCC TTCAGAAATT GATGAAACTG  
1301 AAAGAACGGG CTCTCCGCAT GACGTTCCAC GCAGAAATGG AAGTCCACGT  
1351 GAAGATTCTC CATTGATGAA TGCCTTAGTA GGATGGGCAC ATAAGCACGG  
1401 TGCTAAAACC AAGGAGAGTT CAGAATCAAG TACCCCGGAA ATTTGATTT  
1451 CTGCTCCCAT AGTGAGAGGT TGGAGTCAAG ACAGTTCGGT CAGTTTATTT

-155-

```

1501 GTTATGGAAG ATGATCATAT TTTCTATGAT GTTCCTCGTA GAAAAGATGG
1551 AATCTATGAC GTTCCTAGTT CCCCTAGATG GAGTCCTGCG CGAGAGTTGG
1601 AAGAGGATGT TTTTGGAGAT TATGAAGTTC CTATAACCTC TGCTGAACCA
5 1651 TCTAAAGACA AGAACATCTA CATGACACCT AGATTAGCAA CTCCTGCTAT
1701 CTATGATCTT CCTTCACGTC CAGGATCGTC TGGAGCTCA CGTTCTCCGT
1751 CTTCAGATCG CGTACGAAGC AGCTCACCAA ATAGACGGGG TGTGCTCTT
1801 CCTCCAGTTC CTTACCTGTC TATGAGTGAG GAGGGGAGCA TTTATGAGGA
1851 TATGAGCGGT GCTTCAGGTG CAGGTGAAAG TGATTATGAA GATATGAGCC
10 1901 GTTCCCCCTC TCCTAGAGGC GACTTGGATG AACCCATATA TGCTAATACT
1951 CCTGAAGATA ATCCATTTC TACAGAGAAAT ATAGATAGAA TTTTACAGGA
2001 GAGGTCAGGC GGTGCTTCCG CTTCTCCTGT AGAGCCTATT TATGATGAGA
2051 TCCCATGGAT TCATGGCAGG CCCCTGCTA CACTTCCAAG ACCCGAGAAT
2101 ACATTGACTA ATGTTTCGCT TAGAGTGAGC CCAGGGTTTG GACCAGAAGT
2151 AAGAGCCGCT TTGCTTAGCG AGAGCGTGAG TGCTGTTATG GTCGAAGCAG
15 2201 AGAGTATTGT TCCTCCAACA GAGCCGGGGG ACGGAGAATC AGAATATCTA
2251 GAGCCCTTAG GGGGACTTGT AGCTACAACG AAAATCTTAC TACAAAAAGG
2301 ATGCCTCGT GGAGAGTCGA ATGCTTAG

```

The PSORT algorithm predicts inner membrane (0.3994).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 130A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 130B) and for FACS analysis.

These experiments show that cp6756 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 131

The following *C.pneumoniae* protein (PID 4376761) was expressed <SEQ ID 261; cp6761>:

```

1 MTVAEVKGTG KLVCLGCRVN QYEVQAYRDQ LTILGYQEV L DSEIPADLCI
51 INTCAVTASA ESSGRHAVRQ LCRQNPTAHI VVTGCLGESD KEFFASLDRQ
101 CTLVSNKEKS RLIEKIFSYD TTFPEFKIHS FEGKSRAFIK VQDGCNSFCS
151 YCIIPYLRGR SVSRPAEKIL AEIAGVVDQG YREVVIAGIN VGDYCDGERS
30 201 LASLIEQVDR IPGLIERIRIS SIDPDDITED LHRAITSSRH TCPSSHLVLQ
251 SGSNSILKRM NRKYSRGDFL DCVEKFRASD PRYAFTTDVI VGFPGBSDQD
301 FEDTLRIIED VGFLKVHSFP FSARRRTKAY TFDNQIPNQV IYERKKYLAE
351 VAKRVGQKEM MKRLGETTEV LVEKVTGQVA TGHSPYFEKV SFPVVGTVAI
401 NTLVSVRLDR VEEELIGEI V*

```

The cp6761 nucleotide sequence <SEQ ID 262> is:

```

1 ATGACGGTTG CGGAAGTCAA AGGAACATTT AAGCTGGTCT GTTTAGGCTG
51 TCGGGTGAAT CAGTATGAGG TCCAAGCATA TCGCGACCAG TTGACTATCT
101 TAGGTTACCA AGAGGTCCTG GATTCTGAAA TCCCTGCAGA TTTATGCATA
151 ATCAATACGT GTGCTGTCAC AGCTTCTGCT GAGAGTTCGG GTCGTCATGC
40 201 TGTGCGTCAG TTATGTCGTC AGAACCCTAC AGCACATATT GTTGTCACAG
251 GTTGTTTGGG GGAATCTGAC AAAGAGTTTT TTGCTTCTTT GGATCGGCAA
301 TGCACACTTG TTTCCAATAA AGAAAAATCC CGACTTATAG AAAAAATTTT
351 TTCCTATGAT ACGACCTTCC CTGAGTTCAA GATCCATAGT TTTGAGGGAA
401 AGTCTCGAGC TTTTATTAAA GTTCAAGATG GCTGTAATTC TTTTGTCTCG
45 451 TACTGCATTA TTCCTTATTT GCGGGGCGT TCGGTTTCTC GTCCTGCTGA
501 GAAGATTTTA GCTGAAATCG CAGGGGTGTG AGACCAAGGA TATCGCGAAG
551 TTGTAATTGC AGGAATTAAT GTTGGAGATT ATTGCGATGG AGAGCGTTCA
601 TTGACCTCTT TGATGAACA GGTGGACCGG ATTCTTGGAA TTGAGAGGAT
651 TCGAATTTCC TCTATAGATC CTGATGATAT CACTGAAGAT CTGCACCGTG
50 701 CCATCACCTC ATCGCGTCAC ACTTGTCTTT CGTCACACCT TGTTCTTCAA
751 TCGGGGTCGA ATTCAATTTT AAAGAGAATG AACCGGAAGT ATTCTCGCGG
801 AGATTTTTTA GATTGTGTAG AGAAGTTCCG TGCTTCTGAT CCTCGCTATG
851 CCTTTACTAC AGATGTGATT GTCGGATTTT CTGGAGAGAG TGATCAAGAT
901 TTTGAAGATA CTTTGAGAAT TATTGAAGAT GTAGGCTTTA TTAAGTGCA
55 951 TAGTTTCCCT TTCAGTGCTC GTCGTCGTAC TAAGGCATAT ACTTTTGATA
1001 ATCAGATTCC CAATCAGGTG ATCTATGAGA GGAAGAAGTA TCTTGTCTGAG
1051 GTTGCTAAGA GGGTAGGCCA GAAAGAGATG ATGAAGCGTT TAGGAGAGAC

```

-156-

```

1101 TACAGAGGTG CTTGTTGAGA AAGTAACGGG GCAGGTTGCT ACGGGTCACT
1151 CTCCTTATTT TGAAGAGGTT TCTTCCCTG TTGTAGGAAC GGTAGCTATC
1201 AACACTCTAG TTTCTGTGCG TCTTGATAGG GTAGAGGAAG AAGGGCTGAT
1251 TGGGGAGATT GTATGA

```

5 The PSORT algorithm predicts inner membrane (0.1574).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 131A) and also as a his-tagged product. The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (Figure 131B) and for FACS analysis.

10 These experiments show that cp6761 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 132

The following *C.pneumoniae* protein (PID 4376766) was expressed <SEQ ID 263; cp6766>:

```

15 1 MATSVPTSS TSVGEANSSN ERFTERTSRM YYAALVLGAL SCLIFIAMIV
51 IFPQVGLWAV VLGFLGCLL LSLAIVFAVS GLVLGKLEP SREATPPEIV
101 AQKEWTTQQD VLGNEYWRSE LISLFLRGDL HESLIVDSKD RSLDIDQSLQ
151 NILKLEPLST TSLSLKKDCV HINILHLVR QWNLGVDLS PEVTAHAEKL
201 LLFLIEEQYV SPDILKLIRY GDALQATSPL MDWADSGSFS VDADGVFSCR
251 REECSPEDAL AQFDLLALE NPDRFLKDS FLTYIWSSSF FEKFLHRHLE
301 SLQRKLPETA IDVARYBAQI QFSLRYFQK LDLINAMSLD WGYNCAEGEK
20 351 CYESANQRLD NLFIAFSSSV PAMKRLFDKY GSVVRVDRRQ IREQILSNTE
401 ILENESGFLC SLYEYPLSYL IDWAVLLDCV RGTEISLEDQ ADYTVCLQGL
451 DSMLSQFASR LQSGQKVLNP RDVLSEQAAY MLVHGLAAQG VSFQGLKALM
501 YLTAVPQRMW LGALPLFESF PVFNRMKKEFL GESLGD*

```

The cp6766 nucleotide sequence <SEQ ID 264> is:

```

25 1 ATGGCAACCT CTGTTCTCTGT AACTTCATCT ACTTCTGTAG GAGAGGCTAA
51 CTCCTCCAAC GAAAGATTTA CTGAACGAAC ATCGCGAATG TATTACGCAG
101 CTTTAGTCCT AGGGGCTTTG AGCTGTTTAA TTTTATTGTC TATGATTGTC
151 ATTTCCACAC AGGTCGGATT GTGGGCTGTG GTCCTCGGCT TGCTCTTGG
201 ATGTTTACTT TTAAGCTTAG CTATCGTTT TGCTGTCTCC GGTCTCGTTT
30 251 TAGGCAAGAC TTTAGAACCT AGTCGAGAAG CGACTCCTCC AGAAATGTGT
301 GCGCAAAAGG AGTGGACTAC ACAACAAGAT GTCTTAGGGA ATGAGTATTG
351 GCGTTCGAG TTGATTTCTT TGTTCTTACG AGGGGATCTC CACGAATCTC
401 TGATTTGTTGA TTCTAAGGAT CGATCTTTAG ATATTGATCA GAGTTTACAA
451 AATATATTGA AACTTGAGCC CCTATCTACG ACACTTTCGC TGTTAAAGAA
35 501 AGATTGTGTC CACATCAATA TCATTTTACA TTTAGTGAGA CAGTGGAAC
551 TACTGGGAGT GGATCTTAGT CCTGAAGTCA CTGCGCACGC CGAGGAACCT
601 CTACTCTTTT TGATAGAAGA GCAGTATTAC TCTCTGATA TTTTGAAATT
651 GATTTCGCTAC GGAGATGCTT TACAAGCAAC GTCTCCTTTG ATGGATTGGG
701 CAGATTCAGG TTCCTTTAGT GTAGACGCAG ACGGGGTATT TAGCTGTGCG
40 751 AGAGAAGAAT GTTCTCCTGA GGATGCTTTG GCGCAATTCT ATCTTCTTTT
801 GCGGTTGGAA AATCCCGACA GACGCTTCTT AAAGGATTCT TTTCTTACCT
851 ACATTTGGTC GTCCTCATTT TTTGAGAAGT TTTTACATCG CCATCTAGAG
901 AGCTTGCAAA GAAAGCTCCC AGAGACAGCG ATCGATGTCG CCCGCTATGA
951 AGCACAAATA CAAACATTTT TCTCTCGCTA TTTTCAGAAG CTCGATTTGA
45 1001 TAAACGCAAT GTCTTAGAT TGGGGATATA ACTGTGCTGA GGGAGAAAAA
1051 TGTTATGAGA GCGCAAAATCA AAGATTAGAC AACCTATTTA TTGCTTTTTC
1101 TTCTTCTGTT CCTGCTATGA AGCGGCTCTT TGACAAATAT GGTCTGTGTT
1151 TACGGGTAGA TCGTAGGCAG ATTCTGTAGC AGATTCTTTC GAACACTGAA
1201 ATCTTAGAAA ATGAGTCAGG GTTCTCTGTC AGTTTGTATG AATATCCTTT
50 1251 ATCTTATTG ATAGATTGGG CTGTTTGTCT AGACTGTGTT CGCGGTACCG
1301 AAATCTCTCT AGAAGATCAG CCCGATTACA CCGTTTGTCT GCAAGGCTTG
1351 GATTCTATGT TATCTCAATT TGCGAGTCGT TTACAGTCTG GACAAAAAGT
1401 ATTGAATCCT AGAGATGTTT TAAGTGAACA GGCTGCGGTT ATGCTTGTTC
1451 ATGGCTTGGC AGCACAGGGC GTGTCGTTTC AAGGATTGAA AGCTTTGATG
55 1501 TATTTGACAG CCGTTCCCCA AAGAATGTGG TTAGGAGCAT TGCCTTTATT
1551 TGAATCTTTT CCTGTCTTTA ATCGGATGAA AGAATTTCTT GGGGAATCTC
1601 TGGGAGACTA G

```

-157-

The PSORT algorithm predicts inner membrane (0.6158).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 132A) and also as a his-tagged product. The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (Figure 132B) and for FACS analysis.

- 5 These experiments show that cp6766 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 133

The following *C.pneumoniae* protein (PID 4376804) was expressed <SEQ ID 265; cp6804>:

```

10      1 MSNQLQPCIS LGCVSYINSF PLSLQLIKRN DIRCVLAPPA DLLNLLIEGK
      51 LDVALTSSLG AISENLGYVP GFGIAANQRI LSVNLYAAPT FFNSPQPRIA
     101 ATLEBRSSIG LLKVLCRHLW RIPTPHILRF ITTKVLQPTP ENYDGLLLIG
     151 DAALQHPVLP GFVYDLAGS WYDLTKLPFV FALLHSTSW KEHPLPNLAM
     201 EEALQQFESS PEEVLKEAHQ HTGLPPSLLO EYALCQYRL GEEHYESFEK
     251 FREYYGTLYQ QARL

```

- 15 The cp6804 nucleotide sequence <SEQ ID 266> is:

```

      1 ATGTCTAACC AACTCCAGCC ATGTATAAGC TTAGGCTGCG TAAGTTATAT
     51 TAAATTCCTTT CCGCTGTCCC TACAACTCAT AAAAAGAAAC GATATTCGCT
    101 GTGTTCTTGC TCCCCCTGCA GACCTCCTCA ACTTGCTAAT CGAAGGGAAA
    151 CTCGATGTTG CTTTGACCTC ATCCCTAGGA GCTATCTCTC ATAACTTGGG
    201 GTATGTCCCC GGCTTTGGAA TTGCAGCAA CCAACGTATC CTCAGTGTA
    251 ACCTCTATGC AGCTCCCACT TTCTTTAACT CACCGCAACC TCGGATTGCC
    301 GCAACTTTAG AAAGTCGCTC CTCTATAGGA CTCTTAAAG TGCTTTGTCTG
    351 TCATCTCTGG CGCATCCCAA CTCCTCATAT CCTAAGATTG ATAACTACAA
    401 AAGTACTCAG ACAAAACCCCT GAAAATTATG ATGGCCTCCT CCTAATCGGA
    451 GATGCAGCGC TACAACATCC TGTACTTCCT GGATTGTAA CCTATGACCT
    501 TGCCTCGGGG TGGTATGATC TTACAAAGCT ACCTTTGTGTA TTTGCTCTTC
    551 TTCTACACAG CACCTCTTGG AAAGAACATC CCCTACCCAA CCTTGCATG
    601 GAAGAAGCCC TCCAACAGTT CGAATCTTCA CCCGAAGAAG TCCTTAAAGA
    651 AGCTCATCAA CATACAGGTC TGCCCTCTC TCTTCTTCAA GAATACTATG
    701 CCCTATGCCA GTACCGTCTA GGAGAAGAAC ACTACGAAAG CTTTGAAAAA
    751 TTCCGGGAAT ATTATGGAAC CCTCTACCAA CAAGCCCGAC TGTA

```

The PSORT algorithm predicts inner membrane (0.060).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 133A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 133B) and for FACS analysis.

These experiments show that cp6804 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 134

The following *C.pneumoniae* protein (PID 4376805) was expressed <SEQ ID 267; cp6805>:

```

40      1 MSSLLSCGRI EPTRVTCSLK TYLEDTSQNO LSTRLVASV IFLCALLIIL
     51 VCVALSSLIP SIMALATSFT VMGLILFVMS LLGDVAIISY LTYSTVTSYR
    101 QNKRAFEIHK PARSVYYEGV RHWDLGRSSL GTGEIPIVRT LFSPPQNHGL
    151 NHALAAKIFL FMEHPSPEPP NEPLVDWACL IRDFRPHVSS LCFVIEKQGS
    201 SLRTKEGNTI CEAFRSDYDA HFAMVDCYRL IHSKLIIEKM GLKNIDIIPS
    45      251 VMVREDYPSR PEGEYREGLL RMYGGKGAL*

```

The cp6805 nucleotide sequence <SEQ ID 268> is:

-158-

```

1  ATGTCATCAC TACTGAGCTG CGGAAGAATA GAGCCGACTC GGGTTACCTG
51 TAGCTTAAAG ACGTATCTTG AGGATACGAG TCAGAATCAG TTGAGCACAC
101 GTCTAGTTTCG GGCAAGTGTC ATCTTTTAT GCGCATGTGT GATCATTTTG
5  151 GTTGTGTGG CCCTCTCTAG TTTGATTCCA AGCATTATGG CCTTGGCGAC
201 CTCTTTTACG GTAATGGGGT TAATCTTTT TGTGATGTCA CTTCTTGGTG
251 ACGTTGCAAT TATAAGTTAT CTTACTTATA GCACGTGTAC GAGTTACCGG
301 CAAAATAAGA GAGCTTTTGA GATTACAAG CCCGCTCGCT CCGTTTACTA
351 CGAGGGGGTC CGCCATTGGG ATTTAGGACG ATCATCTTTA GGCACAGGCG
10  401 AGATTCCCTAT AGTAAGGACG TTATTCTCTC CATTCAGAA CCATGGTCTT
451 AACCATGCCT TAGCTGCTAA AATTTTCTTA TTTATGGAGC ATTTCAGCCC
501 TGAGCCACCG AACGAGCCTT TGGTGGATTG GGCCTGTTTG ATTCGGGATT
551 TTAGGCCTCA CGTCAGTTCT TGTGCTTTG TTATTGAAAA ACAAGGGTCA
601 TCCTGAGGA CTAAGGAAGG CAATACGATT TGTGAGGCTT TCCGCTCTGA
651 TTACGACGCC CATTTTGCTA TGGTAGATTG CTACCGGTTG ATCCACTCTA
15  701 AGTTGATTAT AGAGAAAATG GGATTGAAGA ATATCGATAT CATTCCGAGT
751 GTCATGGTTC GTGAAGATTA TCCTAGCCGT CCTGGGGAGG GCTATCGCGA
801 AGGCCTATTA CGTATGTATG GTGGCAAGGG GGCTCTGTGA

```

The PSORT algorithm predicts inner membrane (0.711).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 134A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 134B) and for FACS analysis.

These experiments show that cp6805 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 135

The following *C.pneumoniae* protein (PID 4376813) was expressed <SEQ ID 269; cp6813>:

```

1  MSGPSRTESS QVSVLSYVPR DKEIAPKKQF TIAKISTLAI LASLALGALV
51 AGISLTIVLG NPVFLALLIT TALFSVVTFL VYHQMTSKVS SNWQKVLEQN
101 FKPLGKAWQE KNVDCYSNEM QFYNNHLNPK FKVAIQTDAS QPFQPTFLTG
151 LRVIEKNQST GIIFNPVGPT NLIDNFTATNL STILYSTLKD KSVWDTCKQR
30  201 EGGPAKGEDP FSPTFVRVVK LPNEALDQTF NLNLSAEEK SILPTFLGHV
251 CGPKSEELPN QQEYYRQALL AYENCLKAAI ESHAATVALP LFTSVYEVPP
301 EEILPKEGTF YWDNQTAQFC KRALLDAIQN TALRYPQRS LVLILQDPFNT
351 IESQSRSEE*

```

The cp6813 nucleotide sequence <SEQ ID 270> is:

```

35  1  ATGTCAGGAC CCTCAGTAC TGAGAGCTCT CAAGTTTCTG TACTATCCTA
51  TGTGCTCGG GATAAAGAAA TTGCTCCTAA AAAACAGTTT ACCATAGCAA
101 AAATATCCAC TCTTGCAATC CTAGCTTCTT TAGCTTTAGG AGCTTTGGTG
151 GCTGGAATCT CTTTAACGAT AGTATTAGGG AACCTGTAT TTTTGGCTCT
201 TCTCATTACC ACGGCCCTCT TCTCAGTTGT AACCTTCTTA GTCTACCACC
40  251 AAATGACCTC AAAGGTATCT TCTAACTGGC AGAAAGTTCT AGAGCAAAAC
301 TTCAAGCCTT TGGGAAAAGC GTGGCAAGAA AAAAAAGTAG ACTGCTACTC
351 AAACGAGATG CAATTTTACA ATAATCACCT GAACCTTAAG TTCAAGGTAG
401 CGATACAAAC AGATGCGTCT CAACCATTTT AGCCTACTTT CTTAACTGGA
451 CTTAGAGTGA TCGAAAAAAA TCAATCCACA GGGATCATCT TTAATCCCGT
45  501 AGGCCCAACG AATCTGATCG ACAACACTGC AACGAACCTC TCTACTATCC
551 TTTACTCCAC CCTAAAAGAT AAAAGCGTGT GGGATACATG CAAGCAACGC
601 GAAGGGGGTC CCGCAAAAGG AGAAGACCCC TTTTCCCCTA CCGAAGTGAG
651 AGTAGTAAAA CTTCCAAACG AAGCTCTAGA TCAAAAGTTT AATCTAAATT
701 TAAGCTCTGC AGAAAAGAAA AGTATTCTTC CGACCTTTTT AGGCCACGTA
50  751 TGCGGCCCTA AATCTGAAGA GTTACCAAAT CAGCAAGAAT ATTATCGCCA
801 AGCTTTACTA GCGTACGAGA ACTGCCTTAA AGCAGCTATA GAAAGTCATG
851 CAGCAATCGT TGCTCTTCCT CTCTTTACTT CGGTCTATGA AGTGCCTCCA
901 GAAGAGATTCT TTCTTAAAGA AGGCACTTTC TATTGGGACA ACCAACTCA
951 AGCGTTTTGC AAACGCGCTT TATTGGACGC TATTCAAAAT ACGGCCCTAC
55  1001 GCTATCCTCA AAGATCTTTA CTTGTATATC TCCAAGATCC TTTTAATACT
1051 ATAGAATCAC AAAGTCGTTT TGAGGAGTAA

```

The PSORT algorithm predicts inner membrane (0.4291).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 135A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 135B) and for FACS analysis.

- 5 These experiments show that cp6813 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 136

The following *C.pneumoniae* protein (PID 4376844) was expressed <SEQ ID 271; cp6844>:

```

10      1  MWRVVLRLFI  IFILGRAVFP  LRASESFSWE  TSTCLTVLGI  PFIDIILTNN
      51  EDFVAQCGLQ  IGTISSTNNA  KIKEIFLIYK  EKFPPEASISF  KRKEPLNLSQ
     101  SHLSDLGLIC  MRNGETYABG  MANKENGPAL  KQPKDLRLVL  RCPNQPDLLL
     151  YSEKBAEKGI  ETNTCLCNQG  YTLLDGQLIL  YGDSIEKFLK  ETKRKNNHTL
     201  VDLCDQSQVVT  TFLGRFWSLL  NYVQVLFLE  DSAKILAGIP  DLAQATQLLS
     251  HTVPLLFYIT  NDSIHIEQG  KESSFTYNQD  LTEPILGFLF  GYINRGSMEY
     301  CFNCAQSSLG  ET*
```

The cp6844 nucleotide sequence <SEQ ID 272> is:

```

20      1  ATGTGGCGCG  TTGTCTCAG  ATTCCTTATA  ATTTTATCT  TGGGAAGAGC
      51  CGTCTTCCCT  CTAAGAGCTT  CAGAAAGCTT  CTCTGGGAA  ACATCGACCT
     101  GTTTAACAGT  GCTAGGGATT  CCTTTCATAG  ATATTATCCT  CACAACGAAT
     151  GAGGACTTTG  TTGCCAGTG  CGGCCTGCAA  ATAGGAACCA  TTTCTTCGAC
     201  TAATAACGCA  AAAATAAAAG  AAATTTTTF  GATATATAAG  GAAAAATTTC
     251  CAGAAGCCTC  TATCAGTTTC  AAACGAAAAG  AACCTCTAAA  CCTTTCCCAA
     301  TCCCATCTCT  CCGATTTAGG  TATTTTATGT  ATGCGTAACG  GAGAAACTTA
     351  CGCTGAGGGA  ATGGCAAATA  AAGAAAACGG  ACCCGCTCTA  AAACAACCCA
     401  AGGATCTAAG  ATTAGTTTTA  CGTTGTCTTA  ACCAACCAGA  TACCCTGCTC
     451  TACTCGGAAA  AAGAAGCAGA  AAAGGGCATA  GAAACAAATA  CTTGCCTATG
     501  CAATCAGGGA  TACACACTCC  TGGATGGGCA  ATTGATCTCT  TACGGGGATA
     551  GTATAGAAAA  GTTCTGAAA  GAGACCAAAA  GAAAGAATAA  CCACACGCTT
     601  GTTGATCTTT  GTGACTCACA  AGTCGTGACC  ACGTTCCTCG  GTCGCTTTTG
     651  GTCTCTTCTA  AACTACGTTT  AAGTTCCTTT  CCTATCTGAA  GACTCCGCTA
     701  AAATCTTTCG  GGGCATCCCA  GACCTAGCTC  AAGCTACGCA  ATTGCTTTCC
     751  CACACCGTAC  CTTTGCTTTT  TATTTATACC  AACGATCTTA  TTCACATCAT
     801  AGAACAAGGC  AAAGAAAGTA  GTTTTACCTA  TAACCAAGAT  TTAACAGAGC
     851  CCATTTTAGG  ATTTCTCTTT  GGTTACATAA  ATCGCGGCTC  TATGGAATAC
     901  TGCTTTAATT  GTGCACAGTC  TTCATTAGGA  GAAACCTAA
```

The PSORT algorithm predicts inner membrane (0.1786).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 136A) and also in his-tagged form. The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (Figure 136B) and for FACS analysis.

- 40 These experiments show that cp6844 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 137

The following *C.pneumoniae* protein (PID 4377201) was expressed <SEQ ID 273; cp7201>:

```

45      1  VLVGICPSLY  PEHPRSFFYK  VSGDIGSRFD  DRGFVNSGVE  TLPYSSGSFG
      51  IFWISPTDPT  FNFAIVNIFM  RTAGINEVSR  PMTQDTETSL  IEMRDLSEQQ
     101  EANNNTDSLEQ  EESLMGIVGH  TVGGVSMVT  SSPNIFYRIQ  TLLGLPETLA
     151  EARENPTFPN  STIDSLAEIM  MNLVRISDAV  SIFWIFPIVD  TTYNGVLLAV
```

-160-

```

201 CIGFFGINGI CSTFLMLTNP RSRDRWRNL RIMVLCYRSL GSGMNLFDLS
251 NNVMAARRH VTSTVALYA MVTLEFGWTV IQDALQYGF SVRDAFYRYC
301 LRHRYCLTQR NEDSLQTTGT RFQVTRTHLE DQOMVASILN LSVFGLFFGF
351 VGLMTTFGGL EISPSRWDA ANNRTVGIF*

```

5 The cp7201 nucleotide sequence <SEQ ID 274> is:

```

1 GTGCTCGTTG GTATCTGTCC TTCTCTATAT CCAGAACATC CTCGCTCCTT
51 TTATTATCGT GTTCTCGGAG ATATAGGCTC CCGATTTCGAC GATAGAGGAT
101 TTGTAAACTC TGGAGTCGAA ACCCTGCCAT ACTCTTCAGG CAGCTTTGGG
10 151 ATTTTTTTGA TCTCGTTTAC GGATCCACA TTTAAATTTG CTATCGTAAA
201 TACCTTTATG CGAACTGCAG GGATCAATGA AGTCTCTAGA CCCATGACAC
251 AAGATACAGA AACTTCATTG ATAGAAATGA GAGACCTAAG TGAACAACAA
301 GAAGCGAATA ACACAGATTG TTTAGAGCAA GAAGAGAGCT TAATGGGTAT
351 TGTAGGACAT ACTGTGGGAG GAGTTTCCAT GACCGTGACC TCCAGTCCAA
401 ATATCTTTTA TCGTATACAA ACACTTCTGG GACTGCCAGA GACTCTTGCA
15 451 GAAGCTGAAG AAAATCCTAC CTTCCCAAAT TCTACTATAG ATAGCCTTGC
501 AGAAATAATG ATGAACCTCG TAAGGATCTC TGATGCTGTC TCTATTTTCT
551 GGATTTTTC TATCGTAGAT ACTACATATA ATGGAGTTTT ATTAGCCGTC
601 TGTATCGGCT TCTTCGGAAT CAATGGGATT TGTTCACCGT TCCTTATGCT
651 TACGAATCCA CGCTCTCGTC GAGATAGATG GAGGAATTTA CGCATCATGG
20 701 TCTTTTGCTA TCGTTCTTTG GGAAGCGGAA TGAATCTCTT TGATCTTAGC
751 AATAATGTGC GCATGGCAGC ACGTAGGCAT GTGACATCAT GTACAGTAGC
801 TCTCTATGCT ATGGTCACTC TATTTGGATG GACAGTAGCA ATACAAGATG
851 CTTTGCATA TGGTTTCCCT AGCGTTCGGG ATGCCCTCTA TAGATATTGC
901 TTACGCCACA GATATTGCTT AACTCAAAGA AACGAAGACT CTCTGCAAAC
25 951 TACAGGAACG CGCTTTCAGG TTACCGGTAC ACATCTAGAA GATCAACAGA
1001 TGGTGGCTTC TATTTTGAAT TTGAGTGTTC TTGGGCTCTT TTTTGGATTG
1051 TGAGGGCTAA TGACCAGTTC TGGAGGATTA GAAATCTCAC CATCTTGTCG
1101 GTGGGATGCA GCAAATAACC GAACGGTAGG TATTTTMTAG

```

The PSORT algorithm predicts inner membrane (0.3102).

30 The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 137A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 137B) and for FACS analysis.

These experiments show that cp7201 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### 35 Example 138

The following *C.pneumoniae* protein (PID 4377251) was expressed <SEQ ID 275; cp7251>:

```

1 MAPIHGSNAF VEDILHSHPS PQATYFSSTR AQKLHEFKDR HPVLTRIASV
51 IIRIFKVLIG LIILPLGIYW LCQTLCTNSI LPSKNLLKIF KKQPNKTKLK
101 TNYLHALQDY SSKNRVASM RVPILODNVL IDTLEICLSQ APTNRWMLIS
40 151 LGSDCSLEEI ACKEIFDSWQ RFAKLIGANI LVYNYPGVMS STGSSSLKDL
201 ASAHNICTRY LKDKQGGPGA KEIITYGYSL GGLIQAEALR DQKIVANDDT
251 TWIAVKDRCP LFISPEGFHS CRRIGKLVAR LFGWGTKAVE RSQDLPCLEI
301 FLYPTDSLRR STVRQNKLLA PELTLAHAIK NSPVVQNKRF IEVRLSSDID
351 PIDSKTRVAL ATPILKKS*

```

45 The cp7251 nucleotide sequence <SEQ ID 276> is:

```

1 ATGGCTCCAA TTCACGGAAG TAATGCGTTT GTTGAGGATA TTTTACATTG
51 CCACCCCTCT CCACAAGCGA CTTATTTTTC TTCAACACGC GCCCAAAAAC
101 TTCTATGAGT TAAAGACAGG CATCCCGTGC TTACACGGAT TGCTTCTGTA
151 ATTTATAAAA TTTTAAAGT TCTGATAGGG CTGATCATCC TTCCCTTAGG
50 201 AATCTACTGG CTATGTCAAA CGCTTTGTAC AAACCTCGAT CTCCCTTCCA
251 AGAATTTATT AAAAATTTTC AAGAAGCAAC CCAACACTAA AACCTTAAAA
301 ACTAATTATT TGCATGCTTT GCAAGATTAT TCCTCGAAAA ACCGCGTTGC
351 TTCCATGAGA CGAGTTCCTA TCCTCCAGGA TAATGTCTCT ATCGACACTT
401 TGGAAATATG CCTTTCACAA GCACCTACGA ATCGTTGGAT GCTCATTTCT
55 451 TTAGGAAGTG ACTGTAGCTT GGAAGAAATC GCTTGTAAAG AGATCTTTGA

```

-161-

501 TTCTTGGCAA AGATTTGCCA AGTTGATAGG GGCCAATATA CTCGTTTATA  
 551 ACTACCCCGG AGTCATGTCC AGCAGAGGGA GCAGCAGCCT AAAGGACCTA  
 601 GCATCAGCTC ATAATATTG TACAAGATAC CTTAAAGATA AAGAACAGGG  
 651 CCCTGGAGCA AAAGAAATCA TTACCTATGG GTACTCCCTA GGAGGTTTGA  
 5 701 TACAAGCAGA AGCATTGCCA GACCAGAAGA TTGTTGCAAA CGATGATACT  
 751 ACTTGGATAG CAGTCAAAGA TAGGTGTCCT CTCCTTATAT CTCCAGAAGG  
 801 TTTCCACAGT TGCAGACGCA TAGGAAAGCT AGTAGCTCGT CTTTTTGGCT  
 851 GGGGGACCAA AGCCGTAGAG AGAAGCCAAG ACCTTCCCTG CCTAGAAATT  
 901 TTTCTCTATC CTACGGATT CTTACGAAGA TCAACAGTCA GACAGAACAA  
 10 951 GCTCTTAGCA CCTGAACCTA CTCTCGCTCA TCGGATAAAA AATAGTCCCT  
 1001 ATGTTCAAAA TAAAGAATTT ATAGAAGTAC GATTATCGTC TGATATCGAT  
 1051 CCCATCGACA GCAAAACAAG AGTGGCTCTT GCCACACCAA TTTTGAAAAA  
 1101 GCTCTCTTAG

The PSORT algorithm predicts inner membrane (0.4545).

15 The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 138A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 138B) and for FACS analysis.

These experiments show that cp7251 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

## 20 Example 139

The following *C.pneumoniae* protein (PID 4377288) was expressed <SEQ ID 277; cp7288>:

1 MHMSNPISLF SPAELIAKYN LIPKTSPIYP RRELIILEE NACQTRLTNV  
 51 AQVLHPSSLF SMSKKILNPC GCSGGPLCWV ILNILAFIIT SVLFIIILLPV  
 101 NLIVAGLRLF MPLPPKKIVE DLSEPTTEET NEVIQPFIFA LQALLFEDNK  
 25 151 LRSFKIVEQS VGKAPLPNPF LNRLVAISPQ ESQEBMRKIP DLCSQLKKVL  
 201 KSLGVLTPFW KHMLKYFEBGL KNEHDSNPDK KTFPILIKLL IEALTGKSSL  
 251 PKTPSTKERK QAALFIASSC RTCKPTWGEV ITRSLNRLYS IANEGDNQLL  
 301 IWVQBFKERE LMSIQDGDGA KEYRFAAQQH GERYTEAIEQ VLRNESAAKL  
 351 QWHVINIMKF FHGKNLGLVT EHLQDTLGL TLRTQTTVDTH QGREDDADLSA  
 30 401 ALFLNRYLNS GNQLVNSVFK SMQKADPETK ALIREFALDI LYASLRLPQT  
 451 SAHTEVFSTL LMDPETYEPN KACIAYLLYV LKIIEL\*

The cp7288 nucleotide sequence <SEQ ID 278> is:

1 ATGCATATGT CTAACCCCAT CTCTTTGTTT TCCCCTGCAG AGTTAATAGC  
 51 AAAGTACAAT TTAATTCCAA AAACCTCGCC GATTATCCT CCGAGGACGG  
 35 101 AACCTATTAT CTTGGAAGAA AATGCGTGT CAAACACGCT AACCAACGTG  
 151 GCTCAGGTCC TACATCCTTC TAGCCTATTC AGTATGTCAA AAAAAATACT  
 201 GAATCCCTGC GGGTGCTCTG GTGGTCCCTT ATGTTGGGTG ATTCTCAACA  
 251 TCCTAGCATT TATTATTACT TCAGTACTGT TTATCATTCT TTTACCGGTG  
 301 AATCTCATCG TAGCAGGTCT TCGTCTCTTC ATGCCTCTTC CCCCTAAAAA  
 40 351 AATCGTAGAG GATTTAAGTG AACCTACTAC TGAAGAAACG AATGAGGTCA  
 401 TTCAACCCCT CATTTTCGCT TTGCAAGCGT TGCTTTTGA GGATAACAAA  
 451 CTTCGCTCTT TTAATAATGT TGAACAAAGT GTAGGCAAAG CACCCTTACC  
 501 TAATCCCTTT TTAATAGAC TAGTAGCAAT TTGCGCGCAA GAAAGCCAAG  
 551 AAGCCATGCG GAAGATTCCG GATCTATGCT CACAACGAA AAAAGTATTA  
 45 601 AAGTCTCTAG GCGTGCTAAC TCCAGAATGG AAGCACATGC TGAAGTACTT  
 651 TGAGGGACTG AAAAACGAAC ATGATAGTAA TCCTGATAAA AAGACGTTCC  
 701 CAATATTGAT CAAGCTCCTC ATAGAAGCTC TTAAGTGAAG GTCCCTCTTA  
 751 CCCAAACTCT CTAGTACAAA GGAAAAAATG CAAGCGGCCT TATTATTGTC  
 801 AAGTCTTTCG AAGACTTGTA AGCCGACTTG GGGAGAAGTC ATAACCAGAT  
 50 851 CTCCTAACAG ACTCTATAGT ATAGCTAATG AAGGAGACAA TCAGCTTCTG  
 901 ATTTGGGTTC AAGAGTTTAA AGAAGGAGAG CTGATGTCCA TCCAAGATGG  
 951 TGATGATGCT GAAGAGTATC GGTTCGCGC TCAGCAACAC GGTGAGCGTT  
 1001 ACACAGAGGC AATAGAACAA GTTCTACGAA ACGAGTCAGC AGCCAAACTA  
 1051 CAATGGCATG TGATCAACAC TATGAAATTC TTCCATGGGA AAAATCTCGG  
 55 1101 TCTAGTTACA GAACACCTAC AAGATACTCT CGGCGCCCTA ACTTTACGTC  
 1151 AAATACAGT GGACACACAT CAAGGCAGAG AAGACGCTGA TTTGTACGCT  
 1201 GCTCTTTTCC TAAATAAGTA TTTAAATCTT GGAATCAAC TTGTTAATAG



-162-

```

1251 CGTCTTTAAA TCCATGCAAA AAGCAGATCC AGAAACCAAA GCTTTAATCC
1301 GTGAGTTTGC TCTAGATATA TTATATGCAT CCTTACGGCT TCCTCAAACT
1351 TCCGCTCATA CCGAGGTCTT TTCTACACTC TTAATGGACC CAGAGACCTA
1401 TGAACCTAAT AAAGCTTGTA TCGCCTACTT GCTCTATGTA TTAAAGATCA
1451 TCGAACTATA A

```

The PSORT algorithm predicts inner membrane (0.5989).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 139A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 139B) and for FACS analysis.

- 10 These experiments show that cp7288 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

#### Example 140

The following *C.pneumoniae* protein (PID 4377359) was expressed <SEQ ID 279; cp7359>:

```

1  MPGSVSSPPL SPVIVRERV SSSGSDLIQP HAVLKISILI FALVTILGIV
51  LIVLSSALGA LPSLVLTVSG CIAIAVGLIG LGILVTRLIL STIRKVDAMG
101 YDAAVKEEQY LSRIRELESE NREIRDNRRA VEDQCAHLSE ENKDLRDPEY
151 LHGMTERLIA SLEIENQALV AENILKLDWN ASLSRDFRAY KQKFPLGALE
201 PWKEDIACIM EQNLFLKPEC IAMVKSLEPLE TQRLFLYPKG PQSLVNRFPAP
251 RSRFFQTPKY EYNSRNENED GKVAAVCARL KKEFFSAVLG ACSYEELGGI
301 CERAVALKET LPLPEAVYDT LVQEFNNLLT AESLWKEWCF YSYPYLRPYL
351 SVDYCKRLFV QLFEECLCLKL FTTGSPEDQA LVRLFSYYRN HIPAVLASFG
401 LPPPETGGSV FVLLPKQENL LWSQIEVLAT RYLKDTFVRN SEWTGSFEMM
451 FSYNEMCKEI SEGRIRFAED YETRHSEEFPS PSPLSEEGEG EEFLPPCSEE
501 EVSVLERPDL DVDSMWVWHP PVPKGPL*

```

25 The cp7359 nucleotide sequence <SEQ ID 280> is:

```

1  ATGCCAGGTT CTGTGTCATC ACCTCCTTTG TCTCCTGTAA TGTCCGTGA
51  AAGGGTCCCA TCCTCTTCAG GATCCGACCT CATACAGCCT CATGCTGTTT
101 TAAAGATCTC CATCTAATT TTTGCGCTTG TGACAATTTT AGGAATTGTT
151 CTTGTAGTGT TGTCTAGTGC TTTAGGAGCT CTTCTAGTT TAGTTTTGAC
201 GGTTCCTGGT TGTATGCAA TAGCTGTAGG CCTGATTGGT TTAGGGATTC
251 TTGTGACACG GCTGATTCTC TCTACGATCA GAAAAGTAGA TGCCATGGGT
301 TATGATGCTG CGGTCAAAGA AGAGCAGTAT TTGTACGTA TCAGAGAATT
351 AGAGTCTGAA AATAGAGAGA TTAGAGATAG AAATCGTGCT GTCGAAGATC
401 AGTGTGCCCA TTTATCCGAA GAGAACAAGG ACCTTAGGGA TCCCGAATAT
451 CTACATGGAA TGACTGAAAG GCTCATTCGG AGCTTAGAAA TAGAGAATCA
501 AGCTCTCGTA GCTGAGAACA TTCTTCTCAA AGACTGGAAT GCAAGCCTAT
551 CTAGAGATTT CCGCGCATAT AAGCAAAAAT TTCCTCTTGG GGCATTAGAA
601 CCCTGGAAG AAGATATTGC ATGTATCATG GAACAAAATC TCTTTTAAA
651 ACCGGAATGT ATCGCGATGG TTAAGTCTCT TCCATTAGAG ACGCAACGGC
701 TGTTTTATA TCCAAAAGGA TTTCACTCTT TAGTTAATCG ATTTGCTCCG
751 CGGTCTCGCT TTTTCCAGAC TCCAAAGTAT GAATATAACA GTAGGAATGA
801 AAATGAGGAC GGAAAGGTAG CCGCAGTGTG CGCCCGTTTG AAAAAAGAA
851 TCTTCAGTGC TGTTTTAGGA GCCTGTAGTT ACGAAGAACT AGGGGGCATT
901 TGTGAAAGAG CAGTAGCACT TAAAGAGACG TTGCCATTGC CTGAAGCTGT
951 CTATGATACC CTAGTTCAGG AGTTCCCAA TCTTCTTACT GCTGAGAGTT
1001 TATGAAAGA ATGGTGCTTC TATTCTTATC CCTACCTTCG TCCCTATCTT
1051 TCTGTGGATT ACTGTAAGAG GTTATTGTGA CAACTTTTGG AGGAACCTCTG
1101 CCTAAGCTT TTTACAACGG GATCTCCAGA AGACCAAGCT TTGGTTCCGC
1151 TTTTCTCTTA CTATAGGAAT CATATTCCCG CAGTCTTGGC CTCATTGGGT
1201 TTGCCCCCGC CTGAGACAGG GGGGTCTGTA TTTGTATTGC TACCAAAACA
1251 AGAAAACCTT CTTTGAGATC AAATTGAGGT GCTGGCTACA AGGTATCTCA
1301 AAGATACCTT CGTGAGAAAC TCAGAAATGGA CGGGCTCTTT CGAGATGATG
1351 TTTTCTTATA ACGAGATGTG TAAGGAGATC TCCGAAGGAA GGATTCTGTT
1401 TGCTGAAGAC TATGAAACGA GGCATTCCGA AGAATCCCT CCTTCCCTCTC
1451 TCTCTGAAGA AGGAGAGGGC GAAGAATTCC TTCCTCTTGG CTCTGAAGAA
1501 GAGGTTTCGG TTCTTGAGCG CCCAGATCTA GATGTAGACT CTATGTGGGT
1551 CTGGCATCCG CCGGTCCTA AGGGACCTCT TTAA

```

-163-

The PSORT algorithm predicts inner membrane (0.7453).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 140A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 140B) and for FACS analysis.

- 5 These experiments show that cp7359 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

#### Example 141

The following *C.pneumoniae* protein (PID 4377374) was expressed <SEQ ID 281; cp7374>:

```

10      1 MDKQSSGNSG CIWHPFTQSA LDSTPIKIVR GEGAYLYAES GTRYLDAISS
      51 WWCNLHGHGH PYITKKLCEQ AQKLEHVIFA NPTHEPALEL VSKLAPLLPE
      101 GLERFFFS DN GSTSIBIAMK IAVQYYNQN KAKSHFVGLS NAYHGDTFGA
      151 MSIAGTSPTT VPFHDLFLPS STIAAPYYGK ERLAIAQAKT VFSSESNTAAF
      201 IYEPLIQGAG GMLMYNPEGL KEILKLAKHY GVLCLADEIL TGFGRGTGPIF
      251 ASEFTDIPPD IICLSKGLTG GYLPLALTVT TKEIHDAFVS QDRMKALLHG
15      301 HTFTGNPLGC SAALASLDLT LSPECLQQRQ MIERCHQEFQ EAHGSLWQRC
      351 EVLGTVLALD YPARATGYFS QYRDHLNRFF LERGVLLRPL GNTLYVLPFY
      401 CIQEEDLRIT YSHLQDALCL QPQ*
```

The cp7374 nucleotide sequence <SEQ ID 282> is:

```

20      1 ATGGACAAGC AATCATCAGG GAATTCAGGG TGTATCTGGC ACCCCTTCAC
      51 TCAATCTGCA TTAGATTCTA CACCCATAAA GATTGTAAGG GGAGAAGGTG
      101 CTTACCTCTA TCGCGAATCA GGAACAAGAT ATCTTGATGC GATATCTTCA
      151 TGGTGGTGCA ACCTCCACGG TCGTGGGCAT CCCTACATTA CAAAAAATT
      201 ATGTGAGCAA GCACAGAAGT TAGAACATGT GATCTTCGCA AATTTCACCC
      251 ATGAACCGGC TCTAGAGCTC GTATCGAAAC TCGCTCCCTT CCTTCCTGAA
15      301 GGTCTAGAAC GTTCTCTTTT CTCTGACAAC GGATCAACGT CTATCGAAAT
      351 AGCAATGAAA ATTGCTGTGC AATATTACTA CAATCAAAAC AAGGCTAAGA
      401 GCCATTTTGT TGGACTCAGC AATGCCTATC ACGGAGATAC ATTTGGAGCT
      451 ATGTCGATAG CTGGCAGCAG CCCTACTACA GTTCCCTTTC ATGATCTTTT
      501 TCCTTCCTCC AGTACAATTG CTGCTCCCTA TTATGGCAAG GAAGAGCTTG
30      551 CCATGCCCCA AGCAAAAACA GTCTTTCTG AAAGCAATAT CGCAGCGTTT
      601 ATCTATGAGC CGCTATTGCA AGGTGCTGGA GGGATGTTAA TGTATAATCC
      651 CGAAGGCCCTA AAGGAGATTG TCAAGCTTGC CAAGCATTAC GGGGTTCTCT
      701 GTATTGCTGA TGAATTTCTT ACTGGCTTTG GCCGTACGGG TCCACTGTTT
      751 GCTTCTGAAT TTACAGACAT TCCTCCTGAC ATTATCTGTC TTCTAAAGG
35      801 TCTTACAGGA GGCTATCTCC CTCTAGCCTT GACAGTAACC ACTAAAGAAA
      851 TTCATGATGC CTTTGTCTCC CAAGATCGGA TGAAGGCACT GCTTCATGGC
      901 CATACCTTCA CAGGAAATCC TTTAGGCTGT AGTGTGCCCC TCGCTTCTTT
      951 GGATCTCACC CTATCTCCAG AATGCCTACA ACAAAGGCAA ATGATAGAAC
40      1001 GGTGTCATCA AGAGTTTCAA GAAGCTCATG GTTCCCTATG GCAACGGTGT
      1051 GAGGTTCTGG GCACGGTACT CGCTCTAGAT TACCCTGCAG AAGCTACAGG
      1101 ATATTTTTC AATATAGAG ACCATCTCAA TCGCTTTTTC TTAGAACGTG
      1151 GAGTCCCTCT TCGTCCCTTA GGAACACAC TGTATGTGCT GCCCCCTAC
      1201 TGTATCCAAG AAGAAGATCT CCGGATTATT TATTCTCACC TACAGGATGC
      1251 CCTATGTCTA CAACCACAGT AA
```

- 45 The PSORT algorithm predicts cytoplasm (0.2930).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 141A) and also in his-tagged form. The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (Figure 141B) and for FACS analysis.

- 50 These experiments show that cp7374 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

**Example 142**

The following *C.pneumoniae* protein (PID 4377377) was expressed <SEQ ID 283; cp7377>:

```

1  MREETVSWSL EDIREIVHTP VFELIHKANA ILRSNFLHSE LQTCYLISIK
5  TGGCVEDCAY CAQSSRYHTH VTPEPMMKIV DVVERAKRAV ELGATRVCLG
101 AAWRNAKDDR YFDRVLAMVK SITDLGAEVC CALGMLSEEQ AKKLYDAGLY
151 AYNHNLDDSP EFYETIITTR SYEDRLNLTLD VVNKSGISTC CGGIVGMGES
201 EEDRIKLLHV LATRDHIPES VPVNLWPIID GTPLQDQPII SFWEVLRTIA
251 TARVVFPFSM VRLAAGRAFL TVEQQTLCLF AGANSIFYGD KLLTVENNDI
301 DEDAEMIKLL GLIPRPSFGI ERGNPCYANN S*

```

The cp7377 nucleotide sequence <SEQ ID 284> is:

```

1  ATGCGTGAAG AAACGTGATC CTGGTCATTA GAAGACATCC GCGAAATTTA
51 TCACACTCCC GTATTTGAGC TGATTCACAA AGCCAATGCC ATATTGCGTA
101 GTAATPTCCT CCATTTCAGAA CTGCAGACTT GCTATCTGAT TTCGATTAAA
151 ACTGGTGGAT GCGTTGAAGA TFGCGCCTAC TGTGCCCAAT CTTCCCGCTA
15  201 TCATACCCAC GTCACACCAG AACCTATGAT GAAAAATTGA GACGTTGTGG
251 AAAGGGCAAA ACGTGTCTGA GAGCTAGGCG CCACTCGTGT GTGTCTTGGG
301 GCTGCCCTGGC GCAATGTCAA GGACGATCGA TACTTTGATA GAGTCCCTCGC
351 TATGGTGAAA AGTATCACAG ATCTCGGAGC CGAGGTTTGT TGTGCTTTAG
401 GCATGCTCTC CGAAGAGCAA GCTAAAAAAC TGTATGATGC AGGACTTTAT
20  451 GCCTACAATC ATAATTTAGA CTCTTCTCCG GAATTCATATG AAATAATAAT
501 CACAACACGT TCTTATGAAG ATCGCCTCAA CACTCTTGAT GTAGTAAATA
551 AATCTGGCAT TAGTACATGC TGCGGTGGTA TTGTAGGTAT GGGAGAATCT
601 GAAGAAGACC GTATAAAGCT TCTTCATGTT CTTGCAACAA GAGATCATAT
25  651 CCCAGAATCC GTACCTGTAA ATTTACTTTG GCCGATTGAC GGCACGCCCTT
701 TGCAAGACCA GCCTCCGATT TCTTTCCTGG AAGTCTTGCG AACCATAGCA
751 ACGGCACGGG TTGTTTTCCTC CAGATCCATG GTACGACTTG CTGCAGGACG
801 CGCTTTCCTC ACAGTAGAAC AACAAACCTT ATGTTTCTA GCCGGTGCCA
851 ACTCCATATT CTATGGAGAT AAACGTGTTGA CTGTAGAAAA CAATGATATA
901 GATGAAGATG CTGAAATGAT CAAACTTTTA GGCTTAATCC CTCGCCCTTC
30  951 ATTTGGAATA GAAAGAGGTA ACCCATGTTA TGCCAACAAT TCCTAA

```

The PSORT algorithm predicts cytoplasm (0.2926).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 142A) and also in his-tagged form. The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (Figure 142B) and for FACS analysis.

These experiments show that cp7377 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

**Example 143**

The following *C.pneumoniae* protein (PID 4377407) was expressed <SEQ ID 285; cp7407>:

```

1  MVEPNNSWFR MCGNFNCEWV EVTTTEETTR QSASDISEEA GSSGGAAPIT
40  51 TQPTKITKVE KRVQFNFAQG DESTIHIQIE AGELVDSILS HRRTQGCTEY
101 CYDSYATGCG QRCGSFGRLL CGTYKACCLD REDNQVAGLV HECEQTHGPI
151 AVALAAKTMG LNLMLLVEKN TILSEEQKNE FRQHCSEAKT QLYGTMQSLS
201 QNFFLEGVNS IRERGLDDSL VQAVLSFIAT RSWERTIESE EASGTSSASN
251 STRIPACYIL NTSPLTTSRL SCGSRDARRP SSVGAEPQYV AKKYNDNGMA
45  301 RQLGKIQVTN LKTGDFSALG PFGLLIVKML NSFLLSASQS TSSILKHTGG
351 EICYTCPNFR DIVVLLMLAI GYCANTDET SVVDIHMIDD PIMTIFYRLQ
401 YSYRTGRTSA SFLKKKPSLV RQESLDCPTP AESVPLMSSL EEDENEDDD
451 EDGNLAYQQR ILECSGHLQT LFLGKINKE *

```

The cp7407 nucleotide sequence <SEQ ID 286> is:

```

1  ATGGTTTGCC CAAATAATTC TTGGTTCAGA ATGTGTGGAA ATTTCAACTG
50  51 CGAATGGGTT GAAGTAACAA CAACAGAAGA AACACGCGG CAATCGGCTT
101 CAGATATAAG CGAAGAAGCT GGTTCGAGTG GAGGAGCTGC TCCTATAACT
151 ACGCAACCTA CTAAATTTAC AAAAGTAGAG AACGTGTCC AATTTAATAC

```

-165-

201 TGCTCAAGGT GATGAAAGTA CAATACACAT GATCCAAGAA GCAGGAGAAT  
 251 TGGTAGACTC CATTCATCA CATAGACGAA CGCAAGGATG TACAGAGTAT  
 301 TGTATGACA GTTACGCAAC TGGATGTGGT CAGCGTTGCG GATCTTTTGG  
 351 AAGACTCAAT TGTGGAACGT ATAAAGCGTG TTGCTTAGAC AGAGAGGATA  
 401 ATCAGGTTGC TGGACTTGTC CATGAATGCG AACAGACCCA TGGTCCTATT  
 451 GCCGTTGCTT TAGCTGCTAA AACTATGGGC CTCAACTTAA TGGAACTTGT  
 501 AGAAAAAAC ACTATTTTGT CTGAAGAACA GAAAAATGAA TTTAGACAGC  
 551 ATTGCTCGGA AGCTAAAACC CAACTCTATG GAACGATGCA GAGCCTTTCT  
 601 CAAAACCTTT TCCTTGAAGG AGTCAACAGC ATTAGAGAAC GCCGTCTAGA  
 651 CGATTCACATA GTCCAAGCCG TGCTAAGCTT TATTGCTACA AGGTCTTGGG  
 701 AAAAACTAT AGAATCAGAG GAAGCCTCAG GAACATCTTC TGCTTCTAAT  
 751 TCTACACGCA TTCCTGCGTG CTATATCTTA AATACGAGCC CCTTAACGAC  
 801 GTCACGCCA TCCTGTGGAT CAAGAGATGC GCGACGCCCA TCTTCAGTCG  
 851 GTGCAGAGCC CCAGTACGTA GCAAAAAAAT ACAATGACAA TGGCATGGCC  
 901 AGACAATTAG GAAAAATCCA AGTCACCAAT CTAAAAACAG GAGATTTTTT  
 951 AGCTTTAGGT CCTTTTGGTC TCCTGATTGT GAAATGCTG AATAGCTTTC  
 1001 TCTTATCTGC ATCACAAGC ACATCTTCTA TTCTAAAGCA CACAGGTGGA  
 1051 GAAATATGTT ATACGTGCCC AAATTTTCGT GATATCGTCG TTTTATTGAT  
 1101 GTTAGCGATT GGCTATTGCC CTGCAAATAC CGATGAGACA TCTGTCGTAG  
 1151 ATATACACAT GATAGATGAT CCGATTATGA CCATCTTCTA TCGACTACAA  
 1201 TACAGCTATA GAACAGGGAA AACTTCAGCA TCGTTTTTAA AAAAGAAACC  
 1251 CTCATTAGTA AGACAGGAAA GTCATTGATTG TCCTACCCCT GCAGAATCTG  
 1301 TCCTCTCAT GTCAAGTCTC GAAGAAGAAG ATGAAAATGA AGATGATGAT  
 1351 GAGGATGGGA ATTTGGCGTA TCAACAGCGT ATCCTTGAAT GCTCGGGTCA  
 1401 TTTACAACT CTATTTTCTAG GGATAAAAAT AAACAAAGAA TAA

The PSORT algorithm predicts inner membrane (0.1319).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 143A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 143B) and for FACS analysis.

30 These experiments show that cp7407 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone:

#### Example 144

The following *C.pneumoniae* protein (PID 4376432) was expressed <SEQ ID 287; cp6432>:

35 1 MTRSTIESSD SLCSRSFSQK LSVQTLKNLC ESRLMKITSL VIAFLTLIVG  
 51 GALIALAGGG VLSFPLGLIL GSVLVLFSSI YLVSCCKFFT LKEMTMTCSV  
 101 KSKINIWFEX QRNKDIEKAL ENPDLFGENK RNVGNRSARN QLEMILHETD  
 151 GIILKRYMKG AKMYFYL\*

The cp6432 nucleotide sequence <SEQ ID 288> is:

40 1 ATGACTAGAA GTACTATTGA AAGCAGTGAT TCGCTATGCT CAAGGTCTTT  
 51 TTCTCAAAAA TTAAGTGTC AGACATTAAA AAATCTCTGT GAAAGTAGAT  
 101 TAATGAAGAT CACTTCTCTT GTGATTGCTT TCCTAACTCT AATTGTGGGG  
 151 GGTGCTCTTA TAGCTTTAGC AGGAGGGGGG GTTCTTTCTT TCCCTCTTGG  
 201 GCTAATCTTA GGAAGCGTAC TCGTTTTGTT TTCCTCTATC TATTTAGTCT  
 251 CTTGTTGTAA ATTTTCTACT TTAAAAGAGA TGACAATGAC CTGTAGTGTC  
 45 301 AAATCTAAAA TCAATATATG GTTTGAAAAG CAACGAAACA AAGACATCGA  
 351 AAAGGCATTA GAGAATCCAG ATCTCTTTGG AGAAAATAAG AGAAATGTTG  
 401 GAAATCGTTC GGCAAGAAAT CAACTAGAAA TGATCTTACA CGAGACTGAC  
 451 GGAATTATTT TGAAAAGATA TATGAAAGGA GCTAAAATGT ACTTTTATTT  
 501 ATGA

50 The PSORT algorithm predicts inner membrane (0.5394).

The protein was expressed in *E.coli* and purified as a his-tagged product (Figure 144A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 144B) and for FACS analysis.

-166-

These experiments show that cp6432 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

**Example 145**

The following *C.pneumoniae* protein (PID 4376433) was expressed <SEQ ID 289; cp6433>:

```

5      1  MNWVPKTI DH VDPESIDIR KVVSCYKLIK ECQPEFRSLI SELLGVIRCG
      51  LRLLRKRSKYQ EQARTVSD ED APLFCLTRSY YQDGYLTPLR AGPRDLINHY
     101  IHLRRRENPK HFFSPKHP CY YARLAFNESV CVYRELFDI E RLTKMYVEGD
     151  YSKEQEKNLQ AILSFVKTL D EGKDFLIEHK DTDLIGRGPT DVFCT*

```

The cp6433 nucleotide sequence <SEQ ID 290> is:

```

10      1  ATGAATTGGG TTCCAAAAC AATAGACCAT GTAGATCCAG AATCAGAGAT
      51  AGATATACGT AAAGTCGTCT CCTGCTATAA GTTGATAAAA GAATGTCAAC
     101  CTGAATTCG ATCTCTTATA AGTGAATTAC TAGGAGTGAT TCGGTGTGGC
     151  TTAAGACTAT TAAACGTTT TAAGTATCAA GAACAGGCTA GAACTGTATC
     201  TGATGAAGAT GCACCTCTTT TCTGCCTGAC TCGTTCCTAT TATCAAGATG
15      251  GTTATCTCAC GCCATTAGA GCAGGACCTC GTGATCTTAT AAATCACTAT
      301  ATACACTTGC GTCGCCGAGA GAATCCTAAG CATTTTTCAT GTCTTAAGCA
     351  TCCATGTTAT TATGCTCGAT TGGCTTTTAA TGAGTCAGTG TGTGTCTATA
     401  GAGAACTCTT TGATATAGAG CGACTTACAA AAATGTATGT CGAGGGTGAT
     451  TATTCTAAAG AACAAAGAGAA AAACCTACAG GCTATCTTAA GTTTGTGAA
20      501  AACTCTAGAT GAAGGAAAGG ACTTCTTAT TGAACATAAA GATACCGATC
     551  TCAATTGGGAG AGGTTTACT GATGTGTTCT GCACTTAA

```

The PSORT algorithm predicts cytoplasm (0.4068).

The protein was expressed in *E.coli* and purified as a his-tagged product (Figure 145A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 145B) and for FACS analysis.

These experiments show that cp6433 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

**Example 146**

The following *C.pneumoniae* protein (PID 4376643) was expressed <SEQ ID 291; cp6643>:

```

30      1  MGYLPVSATD VLFESPAAPL INSANTQNRK LIELEKQQA ESSPRITTSV
      51  ILEVLLVIGC CLIVLSLLAI RPAEQFTLET GHPAATAVLA VSGTILLVAV
     101  IILFCFLAAV PFAARKTYKY VRTVDDYASW HSHQQTPTLG TIFSGIVYAE
     151  SQAQL*

```

The cp6643 nucleotide sequence <SEQ ID 292> is:

```

35      1  ATGGGATATC TTCCAGTATC TGCTACGGAC GTTCTTTTGG AAAGTCCAGC
      51  CGCTCCCTTA ATCAATAGCG CAAACACACA AAATCAGAAA CTCATAGAAC
     101  TCAAGGGGAA GCAGCAAGCT GAGTCTTCTC CACGGACAAT CACTTCTGTC
     151  ATATTGGAAG TTCTCCTAGT GATCGGATGC TGCCTCATAG TTCTTAGTTT
     201  ATTGGCAATC CGCCCTGCTC TGCAATTCAC TCTAGAACT GGACATCCAG
40      251  CTGCCATTGC AGTCCTTGCT GTCTCAGGAA CAATTCTATT GGTGGCTGTT
      301  ATCATCTTGT TTTGCTTTCT AGCAGCTGTG CCATTCGCTG CTAAGAAAAC
     351  TTATAAATAT GTTAAGACGG TTGATGACTA TGCTTCTTGG CATTCTCATC
     401  AGCAAACACC GACCCTAGGC ACTATCTTTT CAGGTATCGT CTATGCAGAA
     451  TCCCAGGCGC AATTATAG

```

45 The PSORT algorithm predicts inner membrane (0.6859).

-167-

The protein was expressed in *E.coli* and purified as a his-tagged product (Figure 146A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 146B) and for FACS analysis.

These experiments show that cp6643 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

#### Example 147

The following *C.pneumoniae* protein (PID 4376722) was expressed <SEQ ID 293; cp6722>:

```

1  VSSTLNGVFP SSLPEESADL FITNKEIVAL GEKGNVFLTH SIPMHIAAIT
51  ILVIVALAGI AIICLGCSYQ SILLIAVGIV LTILTLCLQ ALVGFIFIR
101 QLPQQLHTTV QFIREKIRPE SSLQLVTNAQ RKTTQDTLKL YEELCDLSQK
151 EFKLQSTLYQ KRFELSHKNE KTNQN*

```

The cp6722 nucleotide sequence <SEQ ID 294> is:

```

1  GTGTCTAGTA CTTTAAACGG GGTATTTCCTC TCATCCCTTC CGGAAGAGTC
51  TGCTGATTTA TTCATTACGA ATAAGGAGAT CGTAGCTTTG GGGGAGAAGG
101 GCAATGTTTT TCTCACCCAC TCCATTCCCTA TGCATATTGC TCGGATTACG
151 ATCTTAGTGA TTGTAGCTCT TGCTGGAATC GCTATTATCT GTTTGGGTTG
201 CTATAGCCAA AGCATTCTGT TGATTGCCGT TGGCATGTGT CTTACTATTT
251 TGA CTCTTCT CTGCCTACAA GCCTTGGTAG GATTTATTAA ATTCATCCGG
301 CAGCTCCCTC AGCAGCTCCA TACGACAGTA CAATTATCA GGGAGAAGAT
351 TCGACCTGAA TCCTCTCTAC AGCTTGTAAC CAATGCACAG AGAAAAACCA
401 CTC AAGATAC GCTAAAGTTA TACGAAGAAC TCTGCGACCT CTCACAAAAA
451 GAGTTCAAAC TGCAATCAAC TCTTTATCAA AAACGTTTTC AGCTTTCTCA
501 CAAGAATGAA AAGACAAATC AAAACTAG

```

The PSORT algorithm predicts inner membrane (0.6668).

The protein was expressed in *E.coli* and purified as a his-tagged product (Figure 147A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 147B) and for FACS analysis.

These experiments show that cp6722 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

#### Example 148

The following *C.pneumoniae* protein (PID 4377253) was expressed <SEQ ID 295; cp7253>:

```

1  MSELAPCSTG LQVPHTQVH HALDTRRVIL TIAACLSLIA GIVLVGLGAA
51  AILPSLPGVI GGMILILFSS IALIVLYKKT REVDQIALEP LPEMISKDQS
101 IIDFVKTRDY ASLEKKATFA YTHTHYYDGS MVFYREIPRF MLGSYLALRK
151 DMDRQALF*

```

The cp7253 nucleotide sequence <SEQ ID 296> is:

```

1  ATGAGCGAGC TCGCCCCCTG CTCGACAGGA TTGCAGATGG TCCCCCATA
51  GCAGGTCCAT CATGCCCTTG ATACGCGGAG AGTCATTCTA ACGATAGCCG
101 CCTGTCTGTC TTTAATTGCA GGAATCGTGT TGGTTGGCTT AGGTGCTGCA
151 GCAATCCTGC CCTCGCTTTT TGGAGTCATT GGAGGAATGA TTCTTATCT
201 GTTTTCTTCG ATCGCCCTCA TTTATTTATA CAAGAAGACA AGGGAGGTGG
251 ATCAGATTGC TCTGGAGCCT CTTCTTGAGA TGATTCTTAA AGATCAAAGC
301 ATTATAGATT TTGTAAAGAC ACGAGACTAT GCATCTTTAG AAAAGAAAGC
351 GACCTTTGCT TATACTCATA CTCATTATTA CGATGGAAGC ATGGTCTTCT
401 ATAGGGAGAT CCTTAGATTT ATGTTAGGCT CTTATCTCGC GCTTCGCAA
451 GACATGGACC GCCAAGCTCT TTTTGA

```

The PSORT algorithm predicts inner membrane (0.5394).

The protein was expressed in *E.coli* and purified as a his-tagged product (Figure 148A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 148B) and for FACS analysis.

These experiments show that cp7253 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

#### Example 149

The following *C.pneumoniae* protein (PID 4376264) was expressed <SEQ ID 297; cp6264>:

```

1  VISGLLFLLV RREVPVTRSE EIPRGVSVTP SEEPALAKAQ KEPETKKILD
51  RLPKELDQLD TYIQEVFACL ERLKDPKYED RGLLTEAKRK LRVFDVVEKD
101 MMSEFLDIQR VLNERAYYVE HCQDPLENIA YEIPSSQELR DYVCAGVCGY
151 LPSGDARADR LKRSVKVMD RFRVTVKSW EASVMLDHSY GVARELFKKA
201 VGVLEESVYK ILFKSYRDAF YCEKAKIQR DGRFKWL*

```

The cp6264 nucleotide sequence <SEQ ID 298> is:

```

1  GTGATTTCGG GACTTCTATT CCTTCTAGTA AGACGAGAGG TTCCGACAGT
51  ACGTTCAGAG GAAATTCCCA GAGGGGTTTC TGTGACCCCT TCTGAAGAGC
101 CTGCTCTAGA GAAGGCTCAA AAAGAACCGG AGACAAAGAA AATTTTAGAT
151 CGGTTGCCGA AGGAATTGGA TCAGTTAGAT ACGTATATTC AGGAAGTGTT
201 TGCAATGTTA GAGAGGCTGA AGGATCCTAA GTACGAAGAT CGAGGTCCTT
251 TAACAGAGGC GAAGGAGAAA CTTTCGAGTTT TTGACGTTGT TGAGAAAGAT
301 ATGATGTCAG AGTTTCTAGA CATAACAACA GTGTTGAATG AGGAAGCATA
351 TTATGTAGAA CATTGTCAAG ATCCCCTAGA GAATATAGCC TACGAGATTT
401 TCTCTTCCCA AGAGCTTCGT GATTACTACT GTGCAGGGGT GTGTGGGTAT
451 TTGCCTTCTG GGGATGCTCG AGCGGATCGA TTAAGAGAGT CAGTTAAGGA
501 GGTAATGGAT CGCTTTATGA GGGTGACCTG GAAATCTTGG GAGGCATCAG
551 TCATGTTGGA TCATAGCTAT GGGGTAGCGC GAGAGTTATT CAAGAAGGCA
601 GTAGGAGTAC TAGAGGAGAG TGTCTATAAA ATTCTGTTTA AGAGCTATAG
651 AGATCGGTTT TATGAATGTG AGAAGGCAAA GATCCAGAGG GATGGGCGTT
701 TCAAATGGTT ATAG

```

The PSORT algorithm predicts cytoplasm (0.2817).

The protein was expressed in *E.coli* and purified as a his-tagged product (Figure 149A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 149B) and for FACS analysis.

These experiments show that cp6264 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

#### Example 150

The following *C.pneumoniae* protein (PID 4376266) was expressed <SEQ ID 299; cp6266>:

```

1  MLLISGALF LTLGIPGLSA AISFGLGIGL SALGGVLMIS GLLCLLVKRE
51  IPTVRPEEIP EGVSLAPSEE PALQAAQKTL AQLPKELDQL DTDIQEVFAC
101 LRKLKDSKYE SRSFLNDAKK ELRVFDVVE DTLSEIFELR QIVAQEGWDL
151 NFLINGRSL MMTABESLSD LFHVSKRLGY LPSGDVRGEG LKKSAREIVA
201 RLMSLHCEIH KVAVAFDRNS YAMA EKAFK ALGALEESVY RSLTQSYRDK
251 FLESEKAKIP WNGHITWLRD DAKSGCAEKK LGMPRNVGRN LGKQSFQ*

```

The cp6266 nucleotide sequence <SEQ ID 300> is:

```

1  ATGCTCTTAC TGATTTCAGG AGCTCTCTTT CTGACGTTAG GGATTCCAGG
51  ATTGAGTGCA GCAATTTCTT TTGGATTAGG CATCGGTCTC TCCGCATTAG
101 GAGGAGTGCT GATGATTTTC GGAATACTAT GTCTTTTAGT AAAACGAGAG
151 ATTCCGACAG TACGACCAGA AGAAATTCCT GAAGGGGTTT CGCTGGCTCC

```

-169-

201 TTCTGAGGAG CCAGCTCTAC AGGCAGCTCA GAAGACTTTA GCTCAGCTGC  
 251 CTAAGGAATT GGATCAGTTA GATACAGATA TTCAGGAAGT GTTCGCATGT  
 301 TTAAGAAAGC TGAAAGATTG TAAGTATGAA AGTCGAAAGT TTTTAAACGA  
 351 TGCTAAGAAG GAGCTTCGAG TTTTGTGACTT TGTGGTTGAG GATACCCTCT  
 5 401 CCGAGATTTT CGAGTTGCGG CAGATTGTGG CTCAGAGAGG ATGGGATTTA  
 451 AACTTTTGA TCAATGGGGG ACGAAGCCTC ATGATGACTG CAGAATCTGA  
 501 ATCGCTTGAT TTGTTCATG TATCGAAGCG GCTAGGGTAT TTACCTTCTG  
 551 GGGATGTTTCG AGGGGAGGGG TTAAGAAAT CTGCCAAGGA GATAGTCGCT  
 601 CGTTTGATGA GCTTGCATTG CGAGATTCAC AAGGTGGCGG TAGCGTTTGA  
 10 651 TAGGAATTCC TATGCGATGG CAGAAAAGGC GTTTGCGAAA GCGTTGGGAG  
 701 CTTTAGAAGA GAGTGTGTAT CGGAGTCTGA CGCAGAGTTA TAGAGATAAA  
 751 TTTTGGGAGA GCGAGAGGGC GAAGATCCCA TGGAATGGGC ATATAACCTG  
 801 GTTAAGAGAT GATGCGAAGA GTGGGTGTGC TGAAAAGAAG CTCGGGATGC  
 851 CGAGGAACGT TGAAGAAAT TTAGGAAAGC AGTC'TTTTGG GTAG

15 The PSORT algorithm predicts inner membrane (0.3590).

The protein was expressed in *E.coli* and purified as a his-tag product (Figure 150A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 150) and for FACS analysis.

20 These experiments show that cp6266 is a surface-exposed and immunoaccessible protein and that they it is a useful immunogen. These properties are not evident from the sequence alone.

#### Example 151

The following *C.pneumoniae* protein (PID 4376895) was expressed <SEQ ID 301; cp6895>:

1 MKIKKSFQYS LCQAKRFQNM LPNHFDPCIQ PVNLQLKQDR LAYGELIILL  
 25 51 SKYQQRTFSS LLKEETCSLN RAKQHLLYKI LRDFNTMQHL RSLGLNGWGE  
 101 IPMSPL\*

The cp6895 nucleotide sequence <SEQ ID 302> is:

1 ATGAAGATTA AAAAATCTTT TCAATACAGT TTATGCCAAG CAAAGAGATT  
 51 TCAGAACATG CTGCCAAACC ACTTTGATCC ATGTTTGCAG CCAGTGAATT  
 101 TACAACTCAA ACAAGACAGA TTGGCATAAG GGGAGCTCAT CATATTGCTA  
 30 151 TCTAAATATC AACAAAAGAC CTTTTCCTCT TTGTTGAAGG AAGAAACATG  
 201 TTCTCTTAAT CGTGCGAAGC AGCACTTATT GTATAAGATT TTGAGAGATT  
 251 TTAATACTAT GCAGCATCTA AGGTCCCTCG GATTAAATGG TTGGGGAGAG  
 301 ATCCCTATGA GTCCTGCGCT CTAA

The PSORT algorithm predicts cytoplasm (0.3264).

35 The protein was expressed in *E.coli* and purified as a his-tag product (Figure 151A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 151B) and for FACS analysis.

These experiments show that cp6895 is a surface-exposed and immunoaccessible protein and that it is a useful immunogen. These properties are not evident from the sequence alone.

#### 40 Example 152 and Example 153

The following *C.pneumoniae* protein (PID 4376282) was expressed <SEQ ID 303; cp6282>:

1 MSLLNLPSQ DSASEDSTSQ SQIFDPIRNR ELVSTPEEKV RQRLLSFLMH  
 45 51 KLNYPKLLII IEKELKTLFP LLMRKGTLP KRRPDILIT PPTYTDAQGN  
 101 THNLGDPKPL LLIECKALAV NQNALQQLS YNYSIGATCI AMAGKHSQVS  
 151 ALFNPKTQTL DFYPGLPEYS QLLNYFISLN L\*



-170-

The cp6282 nucleotide sequence <SEQ ID 304> is:

```

1  ATGTCCTTAT TGAACCTTCC CTCAGCCAG GATTCTGCAT CTGAGGACTC
51  CACATCGCAA TCTCAAATCT TCGATCCCAT TAGAAATCGG GAGTTAGTTT
101 CTACTCCCGA AGAAAAAGTC CGCCAAAGGT TGCTCTCCTT CCTAATGCAT
151 AAGCTGAAC TACCCTAAGAA ACTCATCATC ATAGAAAAAG AACTCAAAAC
201 TCTTTTTCCT CTGCTTATGC GTAAAGGAAC CCTAATCCCA AAACGCCGCC
251 CAGATATTCT CATCATCACT CCCCCACAT ACACAGACGC ACAGGGAAC
301 ACTCACAACC TAGGCGACCC AAAACCCCTG CTAATTATCG AATGTAAGGC
351 CTTAGCCGTA AACCAAAATG CACTCAAACA ACTCCTTAGC TATAACTACT
10  401 CTATCGGAGC CACCTGCATT GCTATGGCAG GAAACACTC TCAAGTGTCA
451 GCTCTCTTCA ATCCAAAAAC ACAAACTCTT GATTTTATC CTGGCTCCC
501 AGAGTATTCC CAACTCCTAA ACTACTTTAT TTCTTTAAAC TTATAG

```

The PSORT algorithm predicts cytoplasm (0.362).

The following *C.pneumoniae* protein (PID 4377373) was also expressed <SEQ ID 305; cp7373>:

```

15  1  MSTPTVKHFI HTASRWEPVL KEIVASNYWH AQWINTLSFL ENSGAKRISA
51  SEHPTVEVKEE VLKHAABEPR HGHYLTQIS RISETSLPDY TSKNLLGGLL
101 TKYLYLHLLDL RTCRVLENEY SLSGQTLKTA AYILVTYAIE LRASELYPLY
151 HDILKEAQSK ITVRSIILEE QGHLQEMERE LKDLPHGEEL LGYACQFEGE
201 LCLQFVERLE QMIFDPSSTF TRF*

```

20 The cp7373 nucleotide sequence <SEQ ID 306> is:

```

1  ATGCTACAA CCACAGTAA ACACTTTATC CACACAGCCT CTCGTTGGGA
51  GCCCGTTCTC AAAGAGATCG TAGCTTCCAA CTATTGGCAT GCACAATGGA
101 TAAATACCC TCCCTTTTTA GAAATAGTG GAGCAAAAAA AATCTCCGCA
151 AGTGAACATC CTACGGAGGT AAAGGAAGAA GTTTTAAAC ATGCTGCTGA
25  201 AGAATTTCTG CATGGTCACT ATCTAAAAAC TCAGATTTCT AGAATCTCAG
251 AGACTTCTCT CCTTGACTAT ACATCTAAAA ATCTTCTGGG AGGCTTACTT
301 ACAAAATATT ACCTCCATCT TCTAGATTTA AGGACGTGCC GAGTACTGGA
351 AAATGAATAC TCCCTATCGG GACAAACGTT AAAAACTGCA GCGTATATTT
401 TAGTTACCTA CGCAATCGAA CTTCTGCTTT CTGAACTTTA TCCTCTGTAT
30  451 CACGATATTG TGAAAGAAGC TCAAAGTAAA ATAACGGTAA AATCCATTAT
501 CTTAGAAGAG CAAGGCCATC TGCAAGAGAT GGAACGTGAA CTTAAAGATC
551 TCCCCACGGG GGAGGAAGTC TTAGGCTATG CTTGCCAATT CGAAGGGGAG
601 CTTTGCTTGC AGTTTGTAGA GAGATTAGAA CAAATGATCT TCGATCCTTC
651 CTCGACTTTT ACAAGTTCT AG

```

35 The PSORT algorithm predicts cytoplasm (0.1069).

The proteins were expressed in *E.coli* and purified as his-tag products (Figure 152A; 6282 = lanes 8 & 9; 7373 = lanes 2-4). The recombinant proteins were used to immunise mice, whose sera were used in Western blots (Figures 152B & 153) and for FACS analysis.

40 These experiments show that cp6282 & cp7373 are surface-exposed and immunoaccessible proteins and that they are useful immunogens. These properties are not evident from the sequence alone.

Example 154 ,  
 Example 155 ,  
 Example 156 ,  
 Example 157 and  
 45 Example 158

The following *C.pneumoniae* protein (PID 4376412) was expressed <SEQ ID 307; cp6412>:

```

1  MSSSEVVQFT VHGLGFGGLS SKSVVPFKKS LSDAPRVVCS ILVLTGLGA
51  LVCGLAITCW CVPGVILMGG ICAIVLGAIS LALSFLWLWG LFSNCCGSKR
101 VLPGEGLLRD KLLDGGFSRA APSGMGLPGD GSPRASTPSC LEELQARIQA
50  151 VTQAIDQMSD D*

```

The cp6412 nucleotide sequence <SEQ ID 308> is:

-171-

1 ATGAGCAGTT CGGAAGTTGT TTTCCAGACA GTTCATGGCC TTGGCTTTGG  
 51 TGGATTGTCT TCAAAAAGTG TTGTCCCTTT TAAGAAAAGT CTTTCGGATG  
 101 CGCCCCGTGT TGTGTGCTCG ATTTTAGTTT TGACTCTGGG GTTGGGAGCG  
 151 CTTGTTTGTG GTATTGCCAT TACTTGTGG TGTGTCCCGG GAGTTATTTT  
 5 201 AATGGGGGGA ATTTGCGCTA TAGTTTTAGG TGCAATTTCT TTAGCTTTAA  
 251 GTCTATTTTG GTTGTGGGGT TTATTTTCTA ATTGTGTGG TTCTAAGAGA  
 301 GTTTTACCGG GTGAGGGATT GCTACGGGAT AAGCTTTTAG ATGGTGGATT  
 351 TTCAAGAGCG GCACCTTCAG GAATGGGACT TCCGGGTGAT GGATCTCCAA  
 401 GAGCGTCAAC GCCATCTTGC CTAGAGGAAC TTCAAGCAGA GATACAGGCA  
 10 451 GTTACTCAAG CTATCGATCA GATGTCAGAT GATTGA

The PSORT algorithm predicts inner membrane (0.4864).

The following *C.pneumoniae* protein (PID 4376431) was also expressed <SEQ ID 309; cp6431>:

1 LLAGGSLVTT YPKQGRLRS PEQLRVLDL VQSYPNHLHA IELDCGAIPQ  
 51 DLIATYIIT FADFSTYILS LRSYQANSPS DDTWGIWFGS IDDPVQAVIS  
 15 101 FLKDHFALP STLAQDPLLC TNK\*

The cp6431 nucleotide sequence <SEQ ID 310> is:

1 TTGCGAGCAG GAGGTAGTCT TGTTACAACA TACCCTAAGG AAGGTCAGAG  
 51 ATTGCGCTCC CCAGAACAGT TAAGAGTTCT GGATGATTTA GTGCAAAGCT  
 101 ATCCAAATCA CCTACATGCG ATTGAACCTG ATTGTGGTGC AATCCCTCAA  
 20 151 GATTTGATCG GAGCCACCTA TATCATCAGG TTCGCCGATT TTCCACCTA  
 201 TATCTCTCT TTAAGAAGCT ACCAAGCCAA TTCTCCCTCC GATGATACAT  
 251 GGGGGATTG GTTTGATCT ATTGACGATC CTGTTCAAGC AGTCATATCA  
 301 TTTTAAAAG ATCATGGATT TGCTCTTCCC TCGACCTTAG CTCAAGATCC  
 351 TTTGCTTTGT ACTAACAAGT AA

25 The PSORT algorithm predicts cytoplasm (0.2115).

The following *C.pneumoniae* protein (PID 4376443) was also expressed <SEQ ID 311; cp6443>:

1 MIMTTISNSP SPALNPELSL IPPPTLVSSG TQTSLAYTIP AQGRRSTLRI  
 51 ILDIFIILG LATIISTFIV IFFLNGLNLL STPSIISSSC LIIVGLLFLI  
 30 101 MGLYFMISL DQGLVGLLQK ELSQAEEREE EYIQEIEALR GAPRAESPTL  
 151 SPSTWL\*

The cp6443 nucleotide sequence <SEQ ID 312> is:

1 ATGATTATGA CTAATATATC TAACTACCCC TCCCCTGCAT TGAATCCCGA  
 51 ACTTTCCTTT ATTCCTCCAC CAACACTTGT ATCTTCAGGT ACGCAAACAT  
 101 CTCTAGCTTA TACGATCCCC GCACAAGGAC GAAGATCCAC CCTACGTATT  
 35 151 ATATTAGATA TATTCTATTAT CATCTTTGGT TTAGCTACGA TCATTTCTAC  
 201 CTTTATTTGT ATTTCTTTT TAAATGGGCT GAACCTGCTC TCGACCCCAT  
 251 CTATTATCTC TTCGTCATGT TTAATCATTG TTGGATTGCT TTTTGTGATT  
 301 ATGGGGTTAT ATTTCTATGAT CTCGAGTTTG GATCAGGGGC TTGTAGGCCT  
 351 TCTGCAAAAG GAACTCTCTC AAGCCGAAGA AAGAGAAGAA GAGTATATCC  
 40 401 AGGAAATCGA AGCTTTAAGA GGAGCTCCTA GAGCAGAATC TCCACAGAG  
 451 TCTCCTAGTA CCTGGTTATG A

The PSORT algorithm predicts inner membrane (0.5585).

The following *C.pneumoniae* protein (PID 4376496) was also expressed <SEQ ID 313; cp6496>:

1 MLIGRYSSDD QFTEATKNTF TIILGFVVRD NLEGLTNPIS EIVSETSSSI  
 45 51 KDSVLRSLPI LGSILGCARL YSTLSTNDPL DETQEKIWHI IFGALETGLL  
 101 GILILLFKII FVILHCIFHL VIGFCK\*

The cp6496 nucleotide sequence <SEQ ID 314> is:

1 ATGCTAATAG GCAGATACAG TAGTGATGAC CAATTCAGT AAGCAACAAA  
 51 AAACACCCCA ACCATAATTA AGCTAGGTTT TGTAGAGAT AATCTCGAGG  
 50 101 GATTAACGAA CCCTATCTCT GAAATCGTCT CGGAAACCTC CTCTCTATT  
 151 AAAGATTCCG TTCTTCGCTC TCTTCTATT TTAGGTPCCA TTTTAGGATG  
 201 CGCCCGACTT TACAGCACAC TCTCTACAAA TGATCTCTTT GACGAACTC  
 251 AAGAAAAGAT TTGGCACACT ATATTGGAG CCTTAGAAAC CTTAGGCTTA  
 301 GGGATTCTCA TCCTCTTATT TAAAATTATT TTTGTTATAT TAACTGCAT  
 55 351 ATTTTATCTA GTTATTGGGT TCTGCAATA A

-172-

The PSORT algorithm predicts inner membrane (0.5989).

The following *C.pneumoniae* protein (PID 4376654) was also expressed <SEQ ID 315; cp6654>:

```

1  MKTKMNSRKK AGQWAIFNSP TPGVSSTLVL AWTPWGYDYK DVQDILERRK
51 PMSSSLSEKD SKEFLKNLFPV DLENGFTSV HIHAEAFPT LDHTGKPHFK
101 RDNVYLPGLK LGALNEAAVQ ANVSADTQFT LFLTQDECNP FHDKKRG*

```

The cp6654 nucleotide sequence <SEQ ID 316> is:

```

1  ATGAAACTA AAATGAAGTC TAGAAAAAAA GCAGGTCAAT GGGCAATTTT
51 CAATTCCTCA ACTCCTGGTG TCAGTTC AAC TTTAGTTTGA GCATGGACTC
101 CTTGGGGTTA TTACGACAAG GATGTACAAG ATATCTTAGA AAGAAAAGAT
151 CCGATGAGCT CTTGCGCTTTC TGAAAAAGAC TCAAAGGAGT TCTTGAAAAA
201 TCTGTTTGTG GATCTCTTAG AAAATGGCTT CACATCAGTA CATATTCACG
251 CAGAAGAAGC TTTCACCTCT CTTGATCATA CCGGGAAACC TCACCTTAAA
301 AGAGACAATG TGTACTTACC CGGAAAGTTG TTAGGCGCCT TGAATGAGGC
351 TGCGGTACAA GCCAATGTAA GTGCGGATAC TCAATTTACA TTGTCCTTA
15 401 CTCAAGATGA GTGCAATCCT TTTCATGATA AGAAAAGAGG TTAA

```

The PSORT algorithm predicts cytoplasm (0.0730).

The proteins were expressed in *E.coli* and purified as his-tag products (Figure 154A; 6412 = lanes 2-3; 6431 = lanes 11-12; 6443 = lanes 5-6; 6496 = lanes 8-9; 6654 = lane 10; markers in lanes 1, 4, 7). The recombinant proteins were used to immunise mice, whose sera were used in Western blots (Figures 154B, 155, 156, 157 & 158) and for FACS analysis.

These experiments show that cp6412, cp6431, cp6443, cp6496 & cp6654 are surface-exposed and immunoaccessible proteins and that they are useful immunogens. These properties are not evident from their sequences alone.

## Example 159 and Example 160

The following *C.pneumoniae* protein (PID 4376477) was expressed <SEQ ID 317; cp6477>:

```

1  LLKFFLVCEB LCILTVAATH ALLETPLALS FFKELKTKYV YRAKDILQLH
51 NYKGFILNT SPLCS*

```

The cp6477 nucleotide sequence <SEQ ID 318> is:

```

30 1  TTGCTAAAGT TCTTCTAGT ATGTGAAGAG TTATGTATAC TTAAGTTTGC
51 TACACATAGA GCTCTCTTAG AAATCCTTT AGCTCTATCA TTTTPTAAAG
101 AACTTAAGAC AAAATATGTC TACAGGGCGA AAGACATACT ACAACTACAT
151 AACTATAAAG GATTTACTAT CCTTAATACA TCACCGTTAT GTTCTTAA

```

The PSORT algorithm predicts inner membrane (0.128).

The following *C.pneumoniae* protein (PID 4376435) was also expressed <SEQ ID 319; cp6435>:

```

1  LWSHFPRGFF MLPFCPTILL AKPFLNSEN YGLERLAATVD SYFDLQSQI
51 VFLSKQDQGI TVEELSAKDR KFKPGSMNCT LYTEDPILPA HNSFNCSDI
101 QMRTPISPIH *

```

The cp6435 nucleotide sequence <SEQ ID 320> is:

```

40 1  TTGTGGTCGC ATTTCCCAAG AGGATTTTT ATGCTCCCTT TTGCGCTAC
51 CATCCTTCTT GCTAAACCTT TTTTAAATAG CGAGAATTAC GGCTTAGAAC
101 GTTTAGCTGC AACCGTAGAT TCTTATTTTG ATCTGGGACA GTCTCAAATA
151 GTCCTTCCTAA GCAAACAGGA TCAAGGAATC ACTGTGGAAG AATTGAGTGC
201 TAAAGATAGG AAATTCAAGC CAGGCTCTAT GAACTGTACA CTGTACACTG
45 251 AAGATCCTAT CTTACCTGCT CATAATTCCT TTAGTAATTG CTCTGATATT
301 CAAATGCGTA CTCCGATTAG CCCTATACAT TAA

```

The PSORT algorithm predicts periplasmic space (0.4044).

The proteins were expressed in *E.coli* and purified as his-tag products (Figure 159A; 6435 = lanes 2-4; 6477 = lanes 5-7). The recombinant proteins were used to immunise mice, whose sera were used in Western blots (Figures 159B & 160) and for FACS analysis.

- 5 These experiments show that cp6477 & cp6435 are surface-exposed and immunoaccessible proteins and that they are useful immunogens. These properties are not evident from the sequences alone.

**Example 161 and  
Example 162 and  
Example 163**

- 10 The following *C.pneumoniae* protein (PID 4376441) was expressed <SEQ ID 321; cp6441>:

```

1  VEAGANVLVI DTAHAHSGKV FQTVLEIKSQ FPQISLVVGN LVTAAEAASL
51  ABIGVDAVKV GIGPGSICIT RIVSGVGYPQ ITAITNVAKA LKNSAVTVIA
101 DGRIRYSGDV VKALAAGADC VMLGSLLAGT DEAPGDIVSI DEKLFKRYRG
151 MGS LGAMKQG SADRYFQTQG QKKLVP GGVE GLVAYKGSVH DVLYQILGGI
201 RSGMGYVGA E TLKDLKTKAS FVRITESGRA ESHIHNIYKV QPTLNY

```

The cp6441 nucleotide sequence <SEQ ID 322> is:

```

1  GTGGAAGCTG GAGCAAATGT TCTAGTCATT GACACAGCTC ATGCACACTC
51  TAAAGGAGTA TTCCAAACAG TTTTAGAAAT AAAATCCCAG TTCCACAAA
101 TTTCCTTAGT TGTAGGGAAT CTTGTTACAG CTGAAGCCGC AGTTTCCTTA
201 CTGTACAAC T AGAATCGTTT CAGGGGTCGG TTATCCACAA ATTACTGCCA
251 ATGGGATCCT AGCAAAAGCT CTTAAAAACT CTGCCGTGAC TGTAAATGCT
301 GATGGGAGAA TCCGCTATTC TGGAGATGTG GTAAAAGCAT TAGCAGCAGG
351 AGCAGACTGT GTCATGCTAG GAAGTTTGCT TGCAGGGACT GATGAAGCTC
25  401 CTGGGGATAT CGTTTCTATC GATGAGAAGC TTTTAAAG GTACCGCGGC
451 ATGGGATCCT TAGCGCTAT GAAACAAGGA AGTGTGACC GGTATTTTCA
501 AACACAGGGA CAGAAAAGC TGGTTCCTGG GGGAGTTGAA GGACTAGTCG
551 CTTATAAAGG CTCGTCCAC GATGTCTCT ATCAAATTTT AGGAGGAATA
601 CGCTCAGGTA TGGGGTATGT TGGAGCTGAA ACTCTCAAAG ATTTAAAAAC
30  651 TTAGGCTTCC TTTGTTCGAA TTAAGTGAATC TGAAGAGCT GAAAGTCATA
701 TTCAATAATAT TTACAAAGTT CAACCAACCT TAAATATTA A

```

The PSORT algorithm predicts bacterial inner membrane (0.132).

The following *C.pneumoniae* protein (PID 4376748) was also expressed <SEQ ID 323; cp6748>:

```

1  LFSEGTALNL FRIFAPLRNR VTTEYSRARQ PDLHRIAIVY IGVLDSESSK
35  51  ILERLISYMS CIYSBSQMYL RFFMGKNVNO SAVLSKLHVE NLHIRCGFFS
101 EDAVPESEPF DLSIYVHTDR SCPLPTKKRS SSWELQTVEL PESIYPQSEF
151 LLMRPRLMS*

```

The cp6748 nucleotide sequence <SEQ ID 324> is:

```

1  TTGTCTCTG AGGGGACAGC TCTAAATTTA TTTCGTATAT TTGCTCCACT
40  51  ACGCAACCGT GTGACTACAG AATACAGTCG TGCTAGGCAA CCCGACCTAC
101 ATAGAATTGC CATCGTCTAT ATAGGAGTTC TCGATTGAGA AAGTTCCAAG
151 ATCCTAGAGC GGCTAATCTC TTATATGAGT TGTATCTATT CTGAATCGCA
201 AATGTATTTA AGATTCTTTA TGGGCAAGAA TGTAAATCAA AGTGTGTGAC
251 TCTCAAAATT ACATGTAGAA AATCTGCACA TCCGTTGTGG GTTTTTCAGC
45  301 GAGGATGCTG TTCCAGAGAG TGAGCCCTTC GATCTCTCCA TCTACGTGCA
351 CACAGATCGT AGCTGTCTC TCCCTACGAA AAAACGGAGC AGCTCCTGGG
401 AACTCCAAAC TGTAAGACTC CCAGAGTCAA TATATCCACA GTCGGAATTC
451 CTATTGATGA GACCTCGAAT GCTTTCGTAG

```

The PSORT algorithm predicts cytoplasm (0.170).

- 50 The following *C.pneumoniae* protein (PID 4376881) was also expressed <SEQ ID 325; cp6881>:

-174-

1 MRPHRKHVSS KSLALKQSAS THVEITTKAF RLSMPLKQLI LEKSDHLPMP  
 51 ETIRVVLTSH KDKLGTEVHV VASHGKEILQ TKVHNANPYT AVINAFKKIR  
 101 TMANKHSNKR KDRTKHDLGL AAKEERIAIQ EEQEDRLSNE WLPVEGLDAW  
 151 DSLKTLGYVP ASAKKKISKK KMSIRMLSQD BAIRQLESAA ENFLIFLNEQ  
 201 EHKIQCIVKK HDGNYVLIEP SLKPGFCI\*

The cp6881 nucleotide sequence <SEQ ID 326> is:

1 ATGAGACCTC ATCGTAAACA CGTATCATCT AAAAGCTTAG CTTTAAAGCA  
 51 ATCTGCATCA ACTCATGTAG AGATCACAAC AAAAGCCTTT CGTCTCTCTA  
 101 TGCCTCTAAA ACAGCTGATC CTAGAGAAAA GCGACCACCT CCCCCCTATG  
 151 GAAACAATCC GTGTGGTGCT AACCTCTCAT AAAGATAAGC TAGGCACCGA  
 201 GGTGCATGTT GTAGCTTCTC ATGGCAAAGA AATCCTTCAA ACTAAGGTTT  
 251 ATAACGCAAA CCCATACACT GCAGTGATCA ATGCTTTTAA GAAAATCCGC  
 301 ACCATGGCAA ATAAGCACTC CAATAAACGT AAAGACAGGA CAAAACATGA  
 351 TCTAGGTCTT GCAGCAAAAG AAGAACGTAT CGCAATACAG GAAGAACAAG  
 401 AAGATCGCCT TAGCAACGAG TGGCTTCCTG TCGAAGGCCT CGATGCCTGG  
 451 GATTCTCTAA AAACCTCTGG GTATGTTCCC GCATCAGCGA AAAAGAAGAT  
 501 CTCCAAGAAA AAGATGAGCA TTCGTATGCT ATCTCAAGAC GAGGCTATCC  
 551 GCCAGCTAGA GTCTGCCGCA GAAAACCTCC TGATCTTCTT GAACGAGCAA  
 601 GAGCATAAAA TCCAATGCAT TTATAAAAAA CATGACGGCA ACTATGTCCCT  
 651 TATTGAACCT TCCCTCAAGC CAGGATTCTG CATCTGA

The PSORT algorithm predicts cytoplasm (0.249).

The proteins were expressed in *E.coli* and purified as his-tag products (Figure 161A; 6441= lanes 7-9; 6748 = lanes 2-3; 6881 = lanes 4-6). The recombinant protein was used to immunise mice, whose sera were used in Western blots (Figures 161B, 162 & 163) and for FACS analysis.

These experiments show that cp6441, cp6748 & cp6881 are surface-exposed and immunoaccessible proteins and that they are useful immunogens. These properties are not evident from the sequence alone.

#### Example 164 and Example 165 Example 166

The following *C.pneumoniae* protein (PID 4376444) was expressed <SEQ ID 327; cp6444>:

1 MEQPNQVIQD TTTVLYALNS FDPRLSDDTH RLKQSPLEA ENALGEFIEG  
 51 LDTNSFPLEE VAIPILPGYH PKFYLSFIDR DDQGVHYEVL DGVFLKTVAA  
 101 CIIENSFLTD SMSPELLSEV KEALKR\*

The cp6444 nucleotide sequence <SEQ ID 328> is:

1 ATGGAGCAAC CCAATTGTGT GATTTCAGGAT ACTACAACCTG TTTTGTATGC  
 51 CTAAATAGC TTTGATCCTA GACTTAGTGA TGACACTCAC AGACTTGGGA  
 101 AGCAATCACC TCTTGAAGCA GAAAATGCTC TTGGAGAATT TATTGAAGGT  
 151 TTGGATACAA ATAGCTTTCC TTTAGAGGAA GTTGCCATTC CCATCCTGCC  
 201 AGGTATATCAC CCTAAGTTTT ATTTATCTTT CATAGATAGG GACGATCAAG  
 251 GTGTCCACTA TGAAGTTTTA GATGGCGTAT TTTTAAAGAC AGTCGCTGCT  
 301 TGTATPATAG AGAACTCCTT CTTAACTGAT TCTATGAGCC CGGAGCTTCT  
 351 CAGCGAAGTT AAGGAAGCTC TGAAACGATG A

The PSORT algorithm predicts cytoplasm (0.2031).

The following *C.pneumoniae* protein (PID 4376413) was also expressed <SEQ ID 329; cp6413>:

1 MAVQSIKEAV TSAATSVGCV NCSREAI PAF NTEERATSIA RSVIAAIIAV  
 51 VAISLLGLGL VVLAGCCPLG MAAGATTMLL GVALLAWAIL ITLRLNLNIPK  
 101 AEIPSPGNNG EPNERN SATP PLEGGVAGEA GRGGGSPLTQ LDLNSGAGS\*

The cp6413 nucleotide sequence <SEQ ID 330> is:

1 ATGGCTGTTC AATCTATAAA AGAAGCCGTA ACATCAGCCG CAACATCAGT

-175-

```

51 AGGATGTGTA AACTGTTCTA GAGAGGCTAT ACCAGCATTT AATACAGAGG
101 AGAGAGCAAC GAGTATTGCT AGATCTGTTA TAGCAGCTAT CATTGCTGTT
151 GTAGCTATCT CCTTACTCGG ACTAGGTCTT GTAGTTCTTG CTGGTTGCTG
201 TCCTTTAGGA ATGGCTGCGG GTGCTATAAC AATGCTGCTG GGTGTAGCAT
251 TATTAGCTTG GGCAATACTG ATTACTTTGA GACTGCTTAA TATACCTAAG
301 GCTGAAATAC CGAGTCCAGG GAACAACGGT GAGCCTAATG AAAGAAATTC
351 AGCAACTCCT CCTCTAGAGG GTGGTGTTCG AGGAGAAGCC GGTCCGCGCG
401 GGGGGTCACC TTTAACCCAA CTTGATCTCA ATTCAGGGGC GGAAGTTAG

```

The PSORT algorithm predicts inner membrane (0.6180).

10 The following *C.pneumoniae* protein (PID 4377391) was also expressed <SEQ ID 331; cp7391>:

```

1 MMLRVIELPL LPIKQALEKA FVQYNSYKAK LTKVEPCFRE SPAYITSEER
51 LQSLDQTLER AYKEYQKRFG BPSRLESEVS GCREHLREQV KQFETQGLDL
101 IKEELIFVSD VLFRKMVSCL VSTVHVPFME FYYEYFELHR LRLRAQWMAN
151 ABIYSKVRKA FPEMLKETLE KAKAPREKEY WLLCEERKSK EKRLILNKIE
201 AAQQRVKDLE PPPIKETGKQ KRKKEYSFFI RLKS*

```

The cp7391 nucleotide sequence <SEQ ID 332> is:

```

1 ATGATGCTTC GTGTCATAGA GCTTCCACTA CTTCTATATA AGCAAGCGTT
51 GGAGAGGGCT TTTGTACAAT ATAATAGCTA CAAAGCGAAG TTAACCAAGG
101 TAGAACCTTG CTTTAGAGAG AGCCCTGCCT ATATAACTAG CGAAGAGCGA
151 CTCCAGAGTT TGGATCAGAC TTTAGAACGT GCGTACAAAG AGTACCAGAA
201 GAGATTCCAG GAGCCTTCAC GTTTGGAATC GGAAGTAAGT GGATGTAGAG
251 AGCATCTTAG AGAGCAGGTA AAACAATTG AACTCAAGG ACTAGACTTG
301 ATCAAGAAG AGCTTATTTT TGTAGTGAT GTGTTATTCC GAAAAATGGT
351 CAGTTGCTTA GTGTCGACAG TGCATGTTCC CTTTATGGAG TTTTATTATG
401 AGTATTTTGA GTTGCATAGA TTGAGGTTGC GGGCCCAATG GATGGCGAAT
451 GCCGAGATTT ATAGCAAAGT TAGAAAAGCA TTCCAGAGA TGTGAAGGA
501 GACCTTAGAA AAAGCTAAGG CTCCAGAGA AGAAGAGTAT TGGTTACTTT
551 GCGAGGAGAG AAAGAGTAAG GAGAAGCGTT TGATTCTCAA CAAGATAGAG
601 GCAGCTCAGC AGCGGTAAA AGATTTAGAA CCTCCTCCTA TTAAAGAGAC
651 AGGGAAACAG AAACGGAAGA AAGAATATTC GTTTTTCATT CGATTAAAAT
701 CGTGA

```

The PSORT algorithm predicts inner membrane (0.1489).

The proteins were expressed in *E.coli* and purified as his-tag and GST-fusion products (Figure 164A; 6444=lanes 11-12; 7391=lanes 2-3; 6413=lanes 4-6). The recombinant protein was used to immunise mice, whose sera were used in Western blots (Figures 164B, 165 & 166) and for FACS analysis.

These experiments show that cp6444, cp6413 & cp7391 are surface-exposed and immunoaccessible proteins and that they are useful immunogens. These properties are not evident from the sequence alone.

Example 167 ,  
Example 168 ,  
Example 169 and  
Example 170

The following *C.pneumoniae* protein (PID 4376463) was expressed <SEQ ID 333; cp6463>:

```

1 MKKKVTIDEA LKEILRLLEGA ATQKEELCAKL LAQGFATTQS SVSRWLRKIQ
51 AVKVAGERGA RYSLPSSTEK TTRHLVLSI RHNASLIVIR TVPGSASWIA
101 ALLDQGLKDE ILGTLAGDDT IFVTPIDEGR LPLLMVSIAN LLQVFLD*

```

The cp6463 nucleotide sequence <SEQ ID 334> is:

```

1 ATGAAAAAAA AAGTAACAT AGATGAGGCT TTAAAAGAAA TTTTACGTCT
51 TGAAGGAGCG GCAACTCAGG AGGAATTATG TGCAAAACTC TTAGCTCAAG
101 GTTTTGCTAC AACCCAGTCG TCTGTATCTC GTTGGCTACG AAAGATTACG
151 CCGTAAAGG TTGCTGGAGA GCGTGGTGCT CGTTATTCCT TACCCTCTTC

```

-176-

201 AACAGAGAAG ACCACGACCC GTCATTTGGT GCTCTCTATT CGCCATAACG  
 251 CCTCTCTTAT TGTAATTCGT ACGGTTCCCTG GTTCAGCTTC TTGGATCGCT  
 301 GCTTTGTTAG ATCAAGGGCT CAAAGATGAA ATTCCTGGAA CTTTGGCAGG  
 351 AGATGACACG ATTTTGTGCA CTCCTATAGA TGAAGGGAGG CTCCCATTGT  
 5 401 TGATGGTTTC GATTGCAAT TTA CTGCAAG TTTTCTTGA TTAA

The PSORT algorithm predicts inner membrane (0.1510).

The following *C.pneumoniae* protein (PID 4376540) was also expressed <SEQ ID 335; cp6540>:

1 MSQCQSSSTS TWEWMKSFVP NWKNPTPLS PIPSEDEFIL AYEPFVLPKT  
 51 DPENQAQANPP GTSTPNVENG IDDLNPLLQ PNEQNANNP GTSGSNPTSL  
 101 PAPERLPETE ENSQEEOQS QNNEDLIG\*

The cp6540 nucleotide sequence <SEQ ID 336> is:

1 ATGTCTCAAT GTCAGAGTAG CAGTACATCT ACCTGGGAAT GGATGAAATC  
 51 TTTTGTGCCA AACTGGAAGA ATCCAACTCC CCCCTTATCT CCTATACCTT  
 101 CTGAGGACGA ATTTATATTA GCATACGAGC CATTTGTTCT ACCGAAAACA  
 151 GATCCAGAAA ACGCACCAAGC TAATCCTCCA GGCACATCTA CACCGAATGT  
 201 AGAAAACGGG ATCGATGATC TCAACCTCT TCTGGGGCAA CCCAACGAAC  
 251 AAAACAATGC CAACAATCCA GGAACCTCTG GATCTAATCC TACATCTCTA  
 301 CCCGCCCCCG AACGACTCCC TGAACTGAA GAGAACAGCC AAGAAGAAGA  
 351 ACAAGGATCT CAAAATAATG AGGATCTTAT AGGATAA

20 The PSORT algorithm predicts cytoplasm (0.3086).

The following *C.pneumoniae* protein (PID 4376743) was also expressed <SEQ ID 337; cp6743>:

1 LREEGSVSFR EYFRAYMCDK IVAQKNFLFT LDAVIKQAGW RSQEKLNLFY  
 51 VESQALGREI KVSLEEYIQS MVGILGSQRT KKSFKFSVDF TPLEQALQER  
 101 CSSDDDEDAT ATSTATGATA SPTDMHEDE\*

25 The cp6743 nucleotide sequence <SEQ ID 338> is:

1 TTGAGAGAAG AAGGTAGTGT TTCTTTCAGA GAATATTTCA GAGCCTATAT  
 51 GTGTGATAAA ATCGTGGCAC AGAAGAACTT CTTATTTACT TTAGACGCTG  
 101 TAATTAACA GGCCTGTGG AGATCACAAG AGAACTCAA TTTATTTTAT  
 151 GTTGAAAGTC AGGCTTTAGG AAGAGAAATC AAAGTCAGCT TAGAGGAATA  
 301 TATTCAGAGT ATGGTCGGGA TTTTGGGATC TCAGAGAACC AAGAAAAGCT  
 251 TTAAGTTTTT TGTGCACTTT ACCCTTTAG AGCAGGCTCT ACAAGAAAGA  
 301 TGCTCTTCTG ATGATGACGA AGATGCAACA GCAACTTCGA CCGCTACAGG  
 351 GGCAACAGCA TCTCCGACTG ACATGCACGA AGATGAGTAA

The PSORT algorithm predicts cytoplasm (0.2769).

35 The following *C.pneumoniae* protein (PID 4377041) was also expressed <SEQ ID 339; cp7041>:

1 MLMLMMIIG ITGGSGAGKT TLTONIKEIF GEDVSVICQD NYYKDRSHYT  
 51 PEERANLIWD HPDAFDNDLL ISDIKRLKNN EIVQAPVDF VLGNRSKTEI  
 101 ETIYPSKVIL VEGILVFENQ ELRLMDIRI FVDTDADERI LRRMVRDVQE  
 151 QGDSVDCIMS RYLSMVKPMH EKFIETPRKY ADIIVHGNYR QNVVTNLSQ  
 40 201 KIKNHLENAL ESDETYMYVN SK\*

The cp7041 nucleotide sequence <SEQ ID 340> is:

1 ATGTTGATGA TGCTTATGAT GATTATTGGA ATTACAGGAG GTTCTGGAGC  
 51 TGGGAAAACC ACCCTAACCC AAAACATTAA AGAAATTTTC GGTGAGGATG  
 101 TGAGTGTAT CTGCCAAGAT AATTATTACA AAGATAGATC TCATTATACT  
 45 151 CCTGAGAAGC GTGCCAATTT AATTTGGGAT CATCCGGACG CCTTTGATAA  
 201 TGACTTATTA ATTTCAAGCA TAAAACGTCT AAAAAATAAT GAGATTGTCC  
 251 AAGCCCCAGT TTTTGATTTT GTTTTAGGTA ATCGATCTAA AACGGAGATA  
 301 GAAACGATCT ATCCATCTAA AGTTATTCTT GTTGAAGGTA TTCTGGTCTT  
 351 TGAAAATCAA GAACTTAGAG ATCTTATGGA TATTAGGATC TTTGTAGACA  
 50 401 CCGATGCTGA TGAAAGGATA CTACGCCGTA TGGTTCGAGA TGTTCAGAA  
 451 CAAGGAGATA GCGTGGACTG CATCATGTCT CGTTATCTTT CTATGGTAAA  
 501 GCCTATGCAT GAGAAATTTA TAGAGCCGAC TCGGAAATAT GCTGATATCA  
 551 TTGTACATGG AAATTACCGA CAAAACGTAG TAACAAATAT TTTGTACAG  
 601 AAAATTAAAA ATCATTTAGA GAATGCCCTG GAAAGCGATG AGACGTATTA  
 55 651 TATGGTCAAC TCTAAGTAA

-177-

The PSORT algorithm predicts inner membrane (0.1022).

The proteins were expressed in *E.coli* and purified as his-tag products (Figure 167A; 6463 = lanes 2-4; 6540 = lanes 5-7; 6743 = lanes 8-9; 7041 = lanes 10-11). The recombinant proteins were used to immunise mice, whose sera were used in Western blots (Figures 167B, 168, 169 & 170) and for FACS analysis.

These experiments show that cp6463, cp6540, cp6743 & cp7041 are surface-exposed and immunoaccessible proteins and that they are useful immunogens. These properties are not evident from the sequence alone.

#### Example 171 and Example 172 and Example 173

The following *C.pneumoniae* protein (PID 4376632) was expressed <SEQ ID 341; cp6632>:

```

1  VQLFQYMNES GWDWLCDFDS QGEGFQLSRL VGLLHSSWAL YEAKEQFYLP
51  EVSLLTWEEEL IEMQLLSKPT KHGVAKDLN VFRKHFQRFR QYLGSLDLNQ
101 RFENTFLNYP KYHLDR*
```

The cp6632 nucleotide sequence <SEQ ID 342> is:

```

1  GTGCAATTAT TTCAATATAT GAATGAGTCC GGATGGGATT GGCTTTGTGA
51  TTTTGATTCT CAAGGCGAGG GATTCCAGTT ATCAGCTCTG GTTGGGCTGT
101 TACATTCGTC CTGGGCATTA TACGAAGCAA AAGAGCAATT TTACCTTCCT
151 GAGGTTTCTC TATTGACCTG GGAAGAACTG ATAGAAATGC AGTTATTAAG
201 CAAACCAACA AAACACGGGG TTGCAAAAGA TCTTTGTAAT GTATTTGAAA
251 AACACTTTCA AAGGTTTAGA CAGTACCTAG GTTCCTTAGA TCTAAATCAA
301 AGGTTGAAAA ATACCTTCTT GAATTATCCT AAATACCATT TAGATAGGGA
351 GTGA
```

The PSORT algorithm predicts cytoplasm (0.3627).

The following *C.pneumoniae* protein (PID 4376648) was also expressed <SEQ ID 343; cp6648>:

```

1  MPVSSAPLPT SHRPSSGNLG LMEPNKALK ARHQDKTKTK IKLLVKILVA
51  ILVIEVLGII AAFFIPGTPP ICLIIILGGLI LTTVLCVLLL VIKLALVNKT
101 EGTTABEQIK RKLSSKSIS*
```

The cp6648 nucleotide sequence <SEQ ID 344> is:

```

1  ATGCCCGTGT CCTCAGCCCC CCTACCCACA AGCCACCGCC CTTCCTCTGG
51  AAATCTAGGC CTCATGGAAC CAAATTCCAA AGCTCTAAAA GCAAAGCATC
101 AAGATAAAAC GACGAAGACG ATTAACCTTT TAGTTAAAAAT CCTTGTGTCG
151 ATTCTAGTAA TAGAAGTTTT AGGAATAATT GCAGCTTTCT TTATTCTCTG
201 GACTCCTCCC ATCTGCTTGA TTATCCTAGG AGGCCTTATT CTACAAACAG
251 TACTCTGTGT GCTTCTTCTT GTTATAAAGC TTGCCCTTGT AAACAAAACC
301 GAAGGAACAA CTGCTGAACA GCAGATAAAA CGTAAACTCT CTCTCTAAAG
351 TATTCTTTAG
```

The PSORT algorithm predicts inner membrane (0.6074).

The following *C.pneumoniae* protein (PID 4376497) was also expressed <SEQ ID 345; cp6497>:

```

1  MKPNSIIFLE NTKHYDPDIFR BGFVRDRHGL MEASDWLLST BITIIRSILG
51  AIPILGNILG AGRLYSVWYT SDEDWKQV *
```

The cp6497 nucleotide sequence <SEQ ID 346> is:

```

1  ATGAAGCCAA ATAGTATTAT TTTTITAGAA AATACTAAGC ATTATCCCGA
51  CATCTTTTCGA GAAGGATTTG TTCGTGATCG TCATGGACTA ATGGAAGCCT
101 CGGATTGGTT ACTTTCTACG GAAATTACGA TCATTGCTC CATTTCTGGA
151 GCTATCCCTA TTTTAGGAAA TATTCTTGGA GCCGGACGAC TCTATAGCGT
```



201 TTGGTATACA AGTGACGAAG ATTGGAAAAA ACAAGTGGTT TGA

The PSORT algorithm predicts inner membrane (0.145).

The proteins were expressed in *E.coli* and purified as his-tag products (Figure 171A; 6632 = lanes 5-7; 6648 = lanes 8-10; 6497 = lanes 2-4). The recombinant proteins were used to immunise mice, whose sera were used in Western blots (Figures 171B, 172, 173) and for FACS analysis.

These experiments show that cp6632, cp6648 and cp6497 are surface-exposed and immunoaccessible proteins and that they are useful immunogens. These properties are not evident from the sequence alone.

Example 174 ,  
Example 175 ,  
Example 176 ,  
Example 177 and  
Example 178

The following *C.pneumoniae* protein (PID 4377200) was expressed <SEQ ID 347; cp7200>:

1 MPVPIDNSSR NLQEVPESE DLEQHAEESP THQSARSSSL QLSLASSAIS  
51 SRVEQLSSLV LGMENSDFSS LRDVPIPSAI YESSTHTPVP TPLVGVGYN  
101 GSQSGYYDTQ RESLHLSQLL GSRRVEVVYN QGNFMEASLL NLCPRRPRRD  
151 PSPISLALLE LWBAFFLEHP PGSTFNPIFF W\*

The cp7200 nucleotide sequence <SEQ ID 348> is:

1 ATGCCCCGTTT CTATAGATAA TTCCTCTCGC AACCTACAAG AAGTTCAGAA  
51 AAGCCCTAGAA GACCTCGAAC AACACGCAGA AGAATCTCCT ACTCATCAAA  
101 GTGCAGAAAG CAGTCTTTTG CAACTGTCTC TAGCCTCCTC AGCAATTTCT  
151 AGTAGAGTAG AACAACTATC TTCCCTCGTC TTAGGAATGG AAAAATTCAGA  
201 TTTCTCTCTT TTAAGAGACG TTCCTATCTT CTCAGCTATC TACGAATCTT  
251 CAACACACAC ACCTGTCCCC ACTCCTCTAG TTGGCGTGGG ATATATCAAC  
301 GGAAGTCAAT CAGGATACTA CGATACACAA AGAGAATCTC TTCACCTCAG  
351 CCAATTGTTA GGAAGCCGAA GAGTTGAAGT TGCTATAAC CAAGGAACT  
401 TCATGGAGGC CTCTTTGCTA AATCTGTGCC CCAGAAGACC TCGAAGAGAT  
451 CCCTCTCCAA TTTCTTTAGC TCTATTAGAG CTCTGGGAAG CATTTTFTTT  
501 AGAACACCCC CCAGGTAGCA CTTTAAATCC AATATTTT TGGTAA

The PSORT algorithm predicts cytoplasm (0.3672).

The following *C.pneumoniae* protein (PID 4377235) was also expressed <SEQ ID 349; cp7235>:

1 LNFVSTLTGS DFYAPVLEKL EBAFADTGTQ VILFSSSPDF IVHPIAQQLG  
51 ISSWYASCYR DQSAEQTIYK KCLTGDKKAQ ILSYIKKINQ ARSHTFSDHI  
101 LDLPFLMLGE EKTIVVRPQGR LKKMAKKYYW NIV\*

The cp7235 nucleotide sequence <SEQ ID 350> is:

1 TTGAATTTTG TATCGACTCT GACCGGCTCC GATTTTATG CTCCTGTTTT  
51 AGAAAACTA GAAGAAGCTT TTGCAGATAC CACAGGACAG GTGATCCTTT  
101 TTTCTTCTTC TCCAGACTTT ATTGTCCACC CCATAGCGCA GCAACTCGGG  
151 ATTAGTTCCT GGTATGCGTC GTGTTATCGC GATCAGTCTG CAGAACAGAC  
201 GATCTATAAA AAATGTCTTA CAGGGGATAA AAAAGCGCAA ATTTTGAGTT  
251 ATATTAAGAA AATTAATCAA GCAAGAAGCC ATACCTTCTC CGACCATATT  
301 TTAGATCTTC CTTTCTTAT GCTGGGAGAA GAGAAAACCG TCGTTCGCCC  
351 TCAGGGACGA CTCAAGAAAA TGGCAAAAA ATATTACTGG AATATCGTTT  
401 AA

The PSORT algorithm predicts cytoplasm (0.3214).

The following *C.pneumoniae* protein (PID 4377268) was also expressed <SEQ ID 351; cp7268>:

1 MMHRYFIPLL ALLIFSPSLV RAELOPSENK KGGWPTQLSC AEGSQLFCKF

-179-

51 EAAYNNAIRE GKPGLVFFS ERPTPEFADL TNGSFSLSSTP IAKGFNVVVL  
 101 CPGLISPLDF FHKMDPVILY MGSFLEMPPE VEAVSGPRLC YILIDEQGGA  
 151 QCQAVLPLET KN\*

The cp7268 nucleotide sequence <SEQ ID 352> is:

5 1 ATGATGCACC GTTATTTTAT TCCTTTATTA GCACTTCTCA TTTTCTCTCC  
 51 TTCTTTAGTC AGGGCAGAGC TACAACCAAG TGAAAACAGA AAAGGGGGGT  
 101 GGCCATACACA ACTTTCCTGT GCAGAAGGTT CGCAACTCTT CTGTAAATTC  
 151 GAAGCTGCCT ATAATAATGC AATTGAGGAA GGGAAACCTG GGATTTTAGT  
 201 CTTTTTCTCT GAGCGACCCA CACCAGAATT TGCCGACTTA ACGAATGGTT  
 10 251 CATTTTCTCT CTCTACGCCA ATCGCCAAGG GCTTTAATGT CGTTGTGTTA  
 301 TGCCCCGGGC TTATCAGTCC CTTAGACTTT TTCCACAAAA TGGATCCTGT  
 351 GATTCTCTAT ATGGGAAGTT TTCTAGAGAT GTTCCCTGAA GTGGAGGCAG  
 401 TTAGTGGCCC TCGCTTATGT TATATCTTAA TAGATGAACA GGGTGGGGCT  
 451 CAATGTCAGG CTGTCCTGCC TTTAGAAACA AAGAATTAG

15 The PSORT algorithm predicts inner membrane (0.1235).

The following *C.pneumoniae* protein (PID 4377375) was also expressed <SEQ ID 353; cp7375>:

1 MQRILIVGID TGVGKTIVSA ILARALNAEY WKPIQAGNLE NSDSNIVHEL  
 51 SGAYCHPEAY RLHKPLSPHK AAQIDNVISIE ESHICAPKTT SNLIETSGG  
 101 FLSPCTSKRL QGDVFSWWSW SWILVSQAYL GSINHTCLTV EAMRSRNLNI  
 20 151 LGMVVNGYPE DEEHWLTQEI KLPIIGTLAK EKEITKTIIS CYAEQWKEVW  
 201 TSNHQGIQGV SGTPLSLNLH\*

The cp7375 nucleotide sequence <SEQ ID 354> is:

1 ATGCAACGTA TCATCATTTGT AGGAATCGAC ACTGGCGTAG GAAAAACCAT  
 51 TGTCAGTGCT ATCCTTGCTA GAGCACTTAA CGCAGAATAC TGGAAACCTA  
 25 101 TACAAGCAGG GAATCTAGAA AATTCAGATA GCAATATTGT TCATGAGCTA  
 151 TCGGGAGCCT ACTGTCATCC CGAAGCTTAT CGATTGCATA AGCCCTTGTC  
 201 TCCACACAAG GCAGCGCAAA TCGATAATGT AAGTATCGAA GAGAGTCATA  
 251 TTTGTGCGCC AAAACAACCT TCGAATCTGA TTATTGAGAC TTCAGGAGGA  
 301 TTTTATATCC CCTGCACATC AAAAAGACTT CAGGGAGATG TGTTTTCTTC  
 30 351 TTGGTCATGT TCTTGGATT TTAGTGAGCCA AGCATATCTC GGAAGTATCA  
 401 ATCACACCTG TTTAACGGTA GAAGCAATGC GCTCACGAAA CCTCAATATC  
 451 TTAGGTATGG TGGTAAATGG GTATCCAGAG GACGAAGAGC ACTGGGTAAC  
 501 TCAAGAAATC AAGCTTCCTA TAATCGGGAC TCTTGCCAAG GAAAAAGAAA  
 551 TCACAAAGAC AATCATAAGC TGTTATGCCG AACAATGGAA GGAAGTATGG  
 35 601 ACAAGCAATC ATCAGGGAAT TCAGGTGTA TCTGGCACCC CTTCACTCAA  
 651 TCTGCATTAG

The PSORT algorithm predicts cytoplasm (0.0049).

The following *C.pneumoniae* protein (PID 4377388) was also expressed <SEQ ID 355; cp7388>:

1 MOVLLSPQLP PPPQHSVSGI SSPSKLRVLA ITFLVFGMLL LISGALFLTL  
 40 51 GIPGLSAAIS FGLGIGLSAL GGVLMSGLL CLLVKREIPT VRPEEIPGV  
 101 SLAPSEEPAL QAAQTLAQL PKELDQLDID IQEVFACLRK LKDSKYESRS  
 151 FLNDAKKELR VFDFVVEDTL SEIFELRQIV AQEGWDLNFI INGGRSLMMT  
 201 ABESLDDLPH VSKRLGYLPS GDVRGEGLLK SAKETVARLM SLHCEIHKVA  
 251 VAFDRNSYAM AEKAPAKALG ALEESVYRSL TQSYRDKFLE SERAKIPWNG  
 45 301 HITWLRDDAK SGCAEKKLRL AEERWKKFRK AVFWVREDGG PDINNLLGDW  
 351 GTVLDPYRQE RMDIETFHLE YKTTFLKRL HRKCALAKTT FEKKRSKKNL  
 401 QAVEERANARR LKYVRDWDYDQ EFQKAGERLE KLHALYPEVS VSIRENKIQE  
 451 TRSNLEKAYE AIEENYRCCV REQEDYWKKE EKREAEFRER GNKILSPEEL  
 501 ESSLEQFDHG LKNFSEKLME LEGHILKLQK EATAEVENKI LSDAESRLEI  
 50 551 VFEDVKEMPC RIEEIEKTLR MAELPLLPTK KAFERACSOY NSCAEMLEKV  
 601 KPYCKBSLAY VTSKERLVS LDEDLRRAYTE CQKRFQDSDG LSEVVRACRE  
 651 QLRERIQEFB TQGLDLVEKE LLCVSSRLRN TECDCVSGVK KEAPPGKKFY  
 701 AQYYDEIYRV RVQSRWMTMS ERLREGVQAC NKMLKAGLSE EDKVLKEEY  
 751 WLYREERKNK EKRLVGTKIV ATQQRVAAFE SIEVPEIPEA PEEKPSLLDK  
 55 801 ARSLPTREDH T

The cp7388 nucleotide sequence <SEQ ID 356> is:

1 ATGCAAGTAC TTCTATCTCC GCAGCTACCC CCCCCCCCC AACACTCTGT  
 51 AGGGTCGATT TCTTCTCCAT CTAAACTTCG CGTTTATAGCG ATTACTTTTT

-180-

101 TAGTTTTTGG TATGCTCTTA CTGATTTTCAG GAGCTCTCTT TCTGACGTTA  
 151 GGGATTCCAG GATTGAGTGC AGCAATTTCT TTTGGATTAG GCATCGGTCT  
 201 CTCCGCATTA GGAGGAGTGC TGATGATTTC GGGACTACTA TGTC'TTTTAG  
 251 TAAAACGAGA GATTCCGACA GTACGACCAG AAGAAATTC TGAAGGGGTT  
 301 TCGCTGGCTC CTTCTGAGGA GCCAGCTCTA CAGGCAGCTC AGAAGACTTT  
 351 AGCTCAGCTG CCTAAGGAAT TGGATCAGTT AGATACAGAT ATTCAGGAAG  
 401 TGTTTCGCATG TTTAAGAAAG CTGAAAGATT CTAAGTATGA AAGTCGAAGT  
 451 TTTTAAACG ATGCTAAGAA GGAGCTTCGA GTTTTGTGACT TTGTGGTTGA  
 501 GGATACCCCT TCGGAGATT TCGAGTTGCG GCAGATTGTG GCTCAAGAGG  
 551 GATGGGATTT AAAC'TTTTGG ATCAATGGGG GACGAAGCCT CATGATGACT  
 601 GCAGAACTCG AATCGCTTGA TTTGTTTCAT GTATCGAAGC GGCTAGGGTA  
 651 TTTACCTTCT GGGGATGTTT GAGGGGAGGG GTTAAAGAAA TCTGCCAAGG  
 701 AGATAGTCGC TCGTTTGATG AGCTTGCATT GCGAGATTCA CAAGGTGGCG  
 751 GTATCGGTTT ATAGGAATTC CTATGCGATG GCAGAAAAGG CGTTTTCGAA  
 801 AGCGTTGGGA GCTTTAGAAG AGAGTGTGTA TCGGAGTCTG ACGCAGAGTT  
 851 ATAGAGATAA ATTTT'TGGAG AGCGAGAGGG CGAAGATCCC ATGGAATGGG  
 901 CATATAACCT GGTTAAGAGA TGATGCGAAG AGTGGGTGTG CTGAAAAGAA  
 951 GCTTCGGGAT GCCGAGGAAC GTTGGAAGAA ATTTAGGAAA GCAGTCTTTT  
 1001 GGGTAGAAGA AGACGGGGGC TTTGACATCA ATAATCTCCT TGGAGACTGG  
 1051 GGGACAGTGC TTGATCCTTA TAGACAAGAG AGAATGGACG AGATAACGTT  
 1101 CCATGAGTTG TATGAAAAAA CTACGTTT'TT GAAAAGACTG CACAGAAAGT  
 1151 GTGCGTTAGC GAAAACAACC TTTGAAAAGA AGAGATCTAA AAAGAATTTG  
 1201 CAGGCAGTCG AGGAGGCGAA TGCAGTAGG TTGAAATATG TAAGGGATTG  
 1251 GTATGATCAG GAGTTTCAGA AAGCAGGGGA GAGATTAGAG AAATGTCATG  
 1301 CTTTGTATCC TGAGGTTTCA GTCTCTATAA GAGAGAACAA AATACAAGAG  
 1351 ACGCGCTCTA ATTTAGAGAA AGCCTATGAG GCTATCGAAG AGAACTATCG  
 1401 TTGCTGTGTC CGAGAGCAAG AGGACTACTG GAAAGAAGAA GAGAAAAGGG  
 1451 AAGCGGAGTT TAGGGAGAGG GGAACAAGA TTC'TTCTCC TGAGGAGCTG  
 1501 GAAAGTTCTT TGGAGCAATT CGACCATGGT TTGAAAAATT TTTC'TGAGAA  
 1551 ATTAATGGAA TTGGAAGGGC ATATCTTAAA ACTTCAGAAA GAAGCCACAG  
 1601 CAGAGGTGGA GAATAAAATA CTTTCAGATG CAGAGAGCCG CTTGAGATT  
 1651 GTATTTGAAG ATGTCAAGGA GATGCCCTGT CGAATTGAGG AGATAGAGAA  
 1701 GACGCTGCGT ATGGCGGAGC TGCCCTACT TCCTACGAAG AAGGCGTTTG  
 1751 AGAAGGCCCT CTCACAATAT AATAGCTGCG CAGAGATGTT GGAGAAGGTG  
 1801 AAGCCTTACT GCAAGGAGAG CCTCGCCTAT GTGACTAGCA AAGAGCGTTT  
 1851 AGTGAGCTTG GATGAAGATT TACGACGAGC CTACACAGAG TGTCAGAAGA  
 1901 GATTCCAGGG GGATTCCGGT TTGGAGTCGG AAGTAAGAGC CTGTGAGAG  
 1951 CAATGCGGAG AGCGGATCCA AGAGTTTGAA ACTCAAGGGC TGGACTTGGT  
 2001 GGAAAAAGAG TTGCTTTGTG TGAGTAGTAG ATTAAGAAAT ACAGAGTGCG  
 2051 ATTTGTATC TGGTGTTAAG AAAGAAGCAC CTCCTGGTAA GAAGTTTAT  
 2101 GCCCAGTATT ATGATGAGAT TTATCGAGTT AGAGTTCAAT CCCGATGGAT  
 2151 GACGATGTCT GAGAGATTGA GAGAGGGAGT TCAAGCATGC AACAGATGT  
 2201 TGAAGGCAGG CCTAAGCGAA GAAGATAAGG TTC'TTAAAGA AGAAGAGTAT  
 2251 TGGTTGTATC GAGAGGAGAG AAAGAATAAA GAGAAACGTT TGGTTGGTAC  
 2301 TAAGATAGTA GCAACGCAGC AGCGAGTTGC AGCATTTGAA TCCATAGAAG  
 2351 TTCCTGAGAT TCCTGAGGCC CCAGAGGAGA AACCGAGTTT GCTGGATAAA  
 2401 GCGCTTCTT TATTACTCG CGAGGACCAT ACCTAG

The PSORT algorithm predicts inner membrane (0.461).

The proteins were expressed in *E.coli* and purified as his-tag products (Figure 174: 7200=lanes 2-3; 7236=lanes 4-5; 7268=lanes 6-8; 7375=lanes 9-10; 7388=lanes 11-12). The recombinant proteins were used to immunise mice, whose sera were used in Western blots (Figures 174, 175, 176, 177 & 178) and for FACS analysis.

These experiments show that cp7200, cp7235, cp7268, cp7375 & cp7388 are surface-exposed and immunoaccessible proteins and that they are useful immunogens. These properties are not evident from the sequence alone.

### Example 179

The following *C.pneumoniae* protein (PID 4376723) was expressed <SEQ ID 357; cp6723>:

-181-

1 MATSVAPSPV PESSPLSHAT EVLNLPNAYI TQPHPIPAAP WETFRSKLST  
 51 KHTLCFALTL LLTLGGTISA GYAGYTGNWI ICGIGLGIIIV LTLILALLLA  
 101 IPLKNKQTGT KLIDBISQDI SSIGSGFVQR YGLMFSTIKS VHLPELTTQN  
 151 QEKTRILNEI EAKKBSIQNL ELKITECQNK LAQKQPKRKS SQKSFMRSIK  
 5 201 HLSKNPVILF DC\*

The cp6723 nucleotide sequence <SEQ ID 358> is:

1 ATGGCAACTT CCGTAGCCCC ATCACCAGTC CCCGAGAGCA GCCCTCTCTC  
 51 TCATGCTACA GAAGTTCTCA ATCTTCCTAA TGCTTATATT ACGCAGCCTC  
 101 ATCCGATTCC AGCGGCTCCT TGGGAGACCT TTCGCTCCAA ACTTTCCACA  
 151 AAGCATACGC TCTGTTTTGC CTTAACTACTA CTGTTAACCT TAGGGGGAAC  
 201 GATCTCAGCA GGTACGCGAG GATATACTGG AAACCTGGATC ATCTGTGGCA  
 251 TCGGCTTGGG AATTATCGTA CTCACACTGA TTCTTGCTCT TCTTCTAGCA  
 301 ATCCCTCTTA AAAATAAGCA GACAGGAACA AAACCTGATTG ATGAGATATC  
 351 TCAAGACATT TCCTCTATAG GATCAGGATT TGTTTCAGAGA TACGGGTTGA  
 15 401 TGTTCTCTAC AATTAAAAGC GTGCATCTTC CAGAGCTGAC AACACAAAAT  
 451 CAAGAAAAAA CAAGAATTTT AAATGAAATT GAAGCGAAAA AGGAATCGAT  
 501 CCAAAATCTT GAGCTTAAAA TTAGTGAAGT CCAAAACAAG TTAGCACAGA  
 551 AACAGCCGAA ACGGAAATCA TCTCAGAAAT CATTATGCG TAGTATTAAG  
 601 CACCTCTCCA AGAACCTGT AATTTTGTTC GATTGCTGA

20 The PSORT algorithm predicts inner membrane (0.6095).

The protein was expressed in *E.coli* and purified as a his-tag product (Figure 179A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 179B) and for FACS analysis.

25 These experiments show that cp6723 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 180

The following *C.pneumoniae* protein (PID 4376749) was expressed <SEQ ID 359; cp6749>:

1 MSYYFSLWYL KVQHPQAAF DFTRSLCSRI SNFALGVIAL LPIIGQLYVG  
 51 LDWLLSRIKK PEFPDSDVDQI VRVEHVVGHD HRSRVEDILK RQRLSLEPRD  
 30 101 EGKVHGDLPs APFF\*

The cp6749 nucleotide sequence <SEQ ID 360> is:

1 ATGAGTTATT ACTTTTCTCT TTGGTATCTG AAGGTGCAAC AGCACTTTCA  
 51 AGCAGCATTT GATTTTACTC GTCCTCTGTG TTCACGAATT TCTAATTTTG  
 101 CTTTGGGAGT GATTGCATTG CTTCTTATTA TTGGGCAGTT GTATGTAGGG  
 35 151 CTGGACTGGC TCCTCTCTAG GATAAAAAAG CCAGAATTTT CTTCCGATGT  
 201 GGATCAGATC GTGCGAGTAG AACACGTCGT GGGTCACGAC CATAGAAGTC  
 251 GAGTTGAAGA TATTCTAAAG AGACAAAGGC TCTCATTAGA GCCTAGAGAC  
 301 GAGGGGAAGG TTCACGGAGA TCTGCCTTCA GTCCTTTTTT TTTGA

The PSORT algorithm predicts inner membrane (0.2996).

40 The protein was expressed in *E.coli* and purified as a his-tag product (Figure 180A). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 180B) and for FACS analysis.

These experiments show that cp6749 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

Example 181 ,  
 Example 182 ,  
 Example 183 ,  
 Example 184 and  
 Example 185

The following *C.pneumoniae* protein (PID 4376301) was expressed <SEQ ID 361; cp6301>:

```

1  LNQDLQNVYQ ECQKATGLES EVSAYRDHLR EQITEFETQG LDVIKELLF
51  VSSTLKSKLS YDPLIADIPC MKFYEEFYDG IDKARVQSRW LEKSERYRKA
101 KKGFPQEMLKE GLFKEDQALK KAEYRLLREK RMNKEKLLIC NKIEAAQQRV
151 QEFGPSDS*

```

The cp6301 nucleotide sequence <SEQ ID 362> is:

```

1  TTGAATCAGG ATTTACAAAA TGTATACCAA GAGTGCCAGA AGGCTACAGG
51  TTTAGAATCG GAAGTGAGTG CATATAGAGA TCATCTTAGA GAGCAGATCA
101 CAGAGTTTGA AACTCAAGGG CTGGACGTGA TAAAAGAAGA ACTTCTTTT
151 GTGAGTAGTA CTCTCAAAAG TAAATTGAGC TATGATCCAT TAATAGCAGA
201 CATTCCTGT ATGAAGTTT ATGAGGAGTA TTATGATGGC ATTGATAAAG
251 CGAGAGTTCA ATCCCGATGG CTGGAGAAGT CTGAGAGGTA TAGAAAGGCG
301 AAGAAGGGAT TCCAAGAGAT GCTGAAGGAA GGCCTATTCA AAGAAGATCA
351 GGCTTTGAAA AAAGCAGAGT ATAGATTACT TCGAGAGAAG AGAATGAATA
401 AGGAGAAGCT TTTGATTTGC AATAAGATAG AAGCAGCTCA GCAGCGAGTC
451 CAAGAATTG GACCTCGGA TTCATAA

```

The PSORT algorithm predicts cytoplasm (0.4621).

The following *C.pneumoniae* protein (PID 4376558) was also expressed <SEQ ID 363; cp6558>:

```

1  MNIPAPQVPV IDEPVVNNTS SYGLSLKSSL RPITYLILAI LAIATLMSVL
51  YFCGIISVGT FVLGMLIPLS VCSVLCVAYL FYQSSIEKT KVFSITSPSV
101 FFSDEDLNLL LGREEDSVSA IDELLKNFPA DDFRRPKMLP YSNFLDEQGR
151 PNESREEDSH TSKIL*

```

The cp6558 nucleotide sequence <SEQ ID 364> is:

```

1  ATGAACATAC CCGCTCCCCA AGTACCAGTC ATAGATGAGC CTGTAGTGAA
51  CAACACAAGT AGCTATGGTC TTTCATTGAA AAGTAGTTTA AGACCGATTA
101 CTTATTTGAT TTTAGCTATC TTAGCTATAG CCACACTGAT GTCTGTCTC
151 TACTTTTGTG GCATCATTAG TGTGGGACG TTTGTTTGG GCATGCTGAT
201 CCCTCTATCG GTCTGCTCTG TTCTTTGCGT TGCCATTMTA TTCTATCAGC
251 AATCTTCTAT AGAAAAGACT AAGTCTTTT CTATAACCAG TCCTTCAGTA
301 TTTTCTCTG ATGAGGATCT TAATTACTC TTAGGTCGAG AAGAAGATTC
351 AGTGCTGCA ATTGATGAAC TTCTTAAGAA CTTTCCAGCT GATGATTTC
401 GTAGCCGAA GATGCTTCCT TATTCAAATT TTCTAGATGA GCAGGGAAGG
451 CCTAATGAGA GTAGGAAGA AGACTCTCAT ACTTCCAAGA TCTTATAA

```

The PSORT algorithm predicts inner membrane (0.4630).

The following *C.pneumoniae* protein (PID 4376630) was also expressed <SEQ ID 365; cp6630>:

```

1  MSMTIVPHAL FKNHCECHST FPLSSRTIVR IAIASLFCIG ALAALGCLAP
51  PVSIVVGSVL AFIAFVILSL VILALIFGEK KLPPTPRIIP DRFTHVIDEA
101 YGLSISAFVR EQQVTLAEFR QFSTALLCNI SPEEKIKQLP SELRSKVESF
151 GISRLAGDLE KNNWPIFEDL LSQTCPLYWL QKFISAGDPQ VCRDLGVPRE
201 CYGYWLGPL GYSTARATIF CKETHILQQ LTKEDVLLK NKALQEKWDT
251 DEVKAIVERI YTTYTARGTL KTEAGGLTKE TISKELLLLS LHGYSFDQLQ
301 LITQLPRDAW DWLCFVDNST AYNLQLCALV GALSSQNLLD ESSIDFDVNL
351 GLYVIQDLKE AVQAFSASDE PKKELGKFL RHLSSVSKRL ESVLRQGLHR
401 IALEHGNARA RYVDVNFVTG ARIHRKTSIF FKD*

```

The cp6630 nucleotide sequence <SEQ ID 366> is:

```

1  ATGAGCATGA CGATCGTTCC ACATGCTTTA TTTAAAAATC ATTGCGAGTG
51  TCATTCTACC TTTCTTTTGA GTTCAAGGAC TATTGTAAGA ATAGCCATTG
101 CCAGCCTCTT TTGTATAGGT GCATTAGCAG CTTTAGGCTG TTTGGCTCCT
151 CCCGTTCTT ATATTGTTGG GAGTGTTTGA GCTTTATTG CTTTGTTCAT
201 TCTTTCTTTA GTAATTTTAG CTTTGATTTT TGGAGAGAAG AAGCTTCCAC

```

-183-

251 CAACACCAAG AATCATTCCT GATAGATTTA CTCACGTGAT AGATGAAGCT  
 301 TATGGCCTTT CAATCTCTGC ATTTGTAAGA GAACAGCAGG TAACATTAGC  
 351 CGAGTTTAGA CAATTTTCTA CTGCCCTGTT GTGTAACATA TCTCCTGAAG  
 401 AGAAATCAA ACAATTGCCT TCTGAATTGC GAAGTAAAGT AGAGAGTTTT  
 451 GGTATTAGCA GGCTCGCAGG TGATTAGAA AAGAATAATT GGCCAATATT  
 501 TGAAGATCTT TTAAGCCAAA CCTGCCCGTT ATATTGGCTT CAGAAATTTA  
 551 TATCAGCAGG AGATCCACAA GTTTGTAGAG ACCTAGGTGT CCCTAGAGAA  
 601 TGTATGGGT ACTATTGGCT AGGGCCTTTG GGATACAGTA CAGCTAAGGC  
 651 TACAATTTTT TGTAAGAGA CGCATCATAT TCTTCAACAA TTAACGAAAG  
 701 AGGAGCTTCT TTTATTAAAA AACAAGGCTC TTCAAGAGAA ATGGGATACT  
 751 GATGAAGTCA AAGCAATTGT AGAGCGTATC TACACTACCT ATACGGCAGC  
 801 AGGAAGTCTA AAGACCGAAG CAGGGGGACT TACAAAGAG ACAATCAGTA  
 851 AGGAATTGCT ATTGTGAGC TTGCATGGCT ATTCCTTTGA TCAGCTACAG  
 901 CTGATCACTC AACTTCCTAG AGATGCTTGG GATTGGCTGT GTTTTGTAGA  
 951 TAACAGTACC GCATACAACC TTCAGCTTTG TGCTCTTGTG GGAGCTTTGT  
 1001 CATCCCAAAA TCTTCTTGAC GAATCTTCTA TCGATTTTGA TGTAAACCTA  
 1051 GGCCTGTATG TGATTCAAGG TCTAAAAGAA GCTGTTCAGG CATTCTCTGC  
 1101 TTCTGATGAG CCAAGAAAG AACTAGGTAA ATTCCTGTGA AGGCATTGTA  
 1151 GTTCAGTTTC TAAGCGATTA GAGAGTGTAT TAAGACAGGG TCTTCACAGA  
 1201 ATAGCTCTAG AGCATGGAAG TGCCAGAGCT AGGGTTTATG ACGTCAATTT  
 1251 TGTAACAGGA GCTAGAATTC ATAGGAAGAC GAGTATCTTC TTTAAAGACT  
 1301 AA

The PSORT algorithm predicts inner membrane (0.7092).

The following *C.pneumoniae* protein (PID 4376633) was also expressed <SEQ ID 367; cp6633>:

25 1 MVNIQPVYRN TQVNSQATQ FSVCQPALSL IIVSVVAVL AIVALVCSQS  
 51 LLSIELGTAL VLVSLILPAS AMFMIYKMRQ EPKELLIPKK IMELIQEHYP  
 101 SIVVDIFRDQ EVSIYBIHHL ISILNKTNVF DKAPVYLQEK LLQFGIERFK  
 151 DVHPSKLPNF EBILLQHCPL HWLGRLVYPM VSDVTPGTYG YWCGPLGLY  
 201 ENAPSLFERR SLLLLKKISF GEFALLEDGL KKNTWSSSEL VQIRQNLFTF  
 251 YYADKEEVDE ABLNADYBQF DSSLHLIFSH KLS\*

The cp6633 nucleotide sequence <SEQ ID 368> is:

1 ATGGTTAATA TACAGCTGT GTATAGGAAT ACCCAAGTCA ACTATAGTCA  
 51 GGCTACCCAA TTTTCGGTGT GCCAGCCAGC GCTTAGCCTG ATTATCGTTT  
 101 CTGTTGTTGC TGCTGTACTC GCTATTGTAG CTTTGGTATG CAGTCAATCT  
 151 CTTTATATCCA TAGAGTTAGG AACTGCTCTT GTTCTAGTTT CTCTTATCTT  
 201 TTTTGTCTCT GCTATGTTTA TGATTATATA GATGAGACAA GAACCTAAGG  
 251 AGTTGCTGAT CCCTAAGAAA ATCATGGAAC TCATCCAAGA ACATTATCCA  
 301 AGTATTGTTG TTGATTTTAT TAGAGATCAG GAGGTTTCCA TTTATGAGAT  
 351 ACATCACTTG ATCTCTATTC TTAATAAGAC GAATGTTTTC GACAAAGCAC  
 401 CAGTATATTT ACAAGAAAAA CTCTTACAGT TTGGCATTGA GAAGTTCAAA  
 451 GATGTACATC CAAGTAAGCT CCCTAATTTT GAAGAAATTC TTCTACAGCA  
 501 TTGCCCATTG CATTTGGTTG GACGCTCGGT ATATCCCATG GTATCGGATG  
 551 TCACTCCAGG AACCTATGGA TACTATTGGT GTGGTCTTTT AGGACTGTAC  
 601 GAGAACGCTC CCTCTCTTTT TGAACGTCGA TCTCTCTAT TGTAAAGAA  
 651 AATTAGCTTT GGAGAGTTTG CTCTTTTAGA AGATGGTCTC AAGAAAAACA  
 701 CGTGGAGTTC TTCGGAATTC GTTCAATCA GACAAAACCT TTTTACAAGA  
 751 TATTATGCTG ATAAAGAAGA GGTAGATGAA GCAGAGTTAA ACGCTGATTA  
 801 CGAACAGTTT GATTCCCTCC TTCACCTTAT TTTTCTCAC AAGCTCTCTT  
 851 GA

50 The PSORT algorithm predicts inner membrane (0.7283).

The following *C.pneumoniae* protein (PID 4376642) was also expressed <SEQ ID 369; cp6642>:

1 MATISPISLT VDHPLVDTKK KSCSNFDKIQ SRILLITAIF AVLVTIGTLL  
 51 IGLLLNIPVI YFLTGISFIA VVLSNFILYK RATTLLKPRA CGKHKEIKPK  
 101 RVSTNLQYSS ISIAINRSKE NWEHQPKDLQ NLPAPSALLT DNPYEIWKAK  
 151 ISLFSLVSLP PGNPEHLI SASENLGKTL LIEETSQNAP ISSYVDTPS  
 201 PKSLINEAIQ ETRVEINTEL PAGDSGERLY WQPDFRGRVF LPQIPTTPEA  
 251 IYQYYVALYV TYIQTAINTN TQIIQIPLYS LREHLYSREL PPQSRMQQSL  
 301 AMITAVKYMA ELHPEYPLTI ACVERSLAQL PQESIEDLS\*

The cp6642 nucleotide sequence <SEQ ID 370> is:

60 1 ATGGCTACAA TCTCACCCAT ATCTTTAACT GTAGATCATC CCCTAGTAGA

-184-

```

51  CACTAAAAAA AAATCCTGCA GCAACTTTGA TAAGATTCAG TCTCGAATTC
101  TATTGATTAC TGCAATCTTT GCTGTCTTAG TTAATATAGG GACCCCTACTT
151  ATTGGGTTGC TTTTAAATAT TCCTGTTATC TATTTCTCTCA CAGGAATTTTC
201  ATTTATTGCT GTTGTCTTGA GCAACTTTAT CCTTTATAAA CGAGCAACCA
5  251  CCCTCTTAAA ACCGCGTGCT TGTGGCAAAC ACAAAGAAAT AAAACCAAAA
301  AGGGTCTCCA CCAACCTACA GTATTCTTCT ATCTCTATCG CAATCAATCG
351  TTCTAAAGAA AACTGGGAAC ACCAACCCTA GGACCTACAG AATCTCCCCG
401  CACCCTCTGC ATTACTCACA GATAACCCCT ACGAGATATG GAAAGCTAAA
451  CATTCAGTGT TTTCCCTAGT ATCCCTCCTA CCGGGAGGCA ATCCAGAACA
10 501  TCTCTTAATT TCAGCTTCCG AAAATTTAGG AAAGACTCTG TTAATTGAAG
551  AAACCTCGCA AAATGCGCCT ATATCTCCTT ACGTAGATAC CACTCCCTCC
601  CCAAAATCCT TGCTCAATGA GGCAATTCAG GAAACCAGGG TAGAAATAAA
651  TACAGAACTC CCTGCGGGAG ATTCAAGAGA ACGTTTATAC TGGCAACCCG
701  ATTTCCGAGG CCGCGTCTTC CTCCACAAA TACCAACAAC TCCTGAAGCC
15 751  ATCTACCAAT ACTACTATGC ACTCTATGTC ACTTATATCC AGACTGCGAT
801  CAATACGAAC ACCCAAATTA TCCAAATCCC TTTATACAGC TTGAGGGAGC
851  ATCTCTATTC TAGAGAATTG CCCCAGCAAT CAAGAATGCA ACAATCTTTG
901  GCTATGATTA CAGCAGTAAA ATACATGGCC GAGCTGCACC CAGAATATCC
951  GCTAACTATT GCTTGTGTTG AAAGATCCTT AGCCCACTA CCTCAAGAAA
20 1001 GTATTGAGGA TCTCTCTTAG

```

The PSORT algorithm predicts inner membrane (0.5288).

The proteins were expressed in *E.coli* and purified as GST-fusion products. The recombinant proteins were used to immunise mice, whose sera were used in Western blots (Figures 181-185) and for FACS analysis.

25 These experiments show that cp6301, cp6558, cp6630, cp6633 and cp6642 are surface-exposed and immunoaccessible proteins, and that they are useful immunogens. These properties are not evident from their sequences alone.

#### Example 186

The following *C.pneumoniae* protein (PID 4376389) was expressed <SEQ ID 371; cp6389>:

```

30 1  MSEVKPLFLK NDSFDLATQR FQNLINMLQE QAEIYNEYEE KNARVQNEIK
51  EQKDFVKRCI EDFEARGLGV LKEELASLIR DFHDKAKAET SMLIEPCIG
101 FYYSIHQEEQ RQRQERLQKM AERYRDCKQV LEAVQVEQKD MISSRVVVDD
151 SYFEEKEEQ KVDNRKKEQD *

```

The cp6389 nucleotide sequence <SEQ ID 372> is:

```

35 1  ATGTCAGAAG TGAAGCCTTT GTTTTAAAG AATGACTCTT TTGATTGGC
51  AACTCAGAGA TTCCAGAATC TAATTAACAT GCTACAAGAG CAAGCCGAGA
101 TATATAACGA GTATGAAGAA AAGAATGCTA GGGTTAGAA TGAGATTAAG
151 GAGCAAAAGG ACTTTGTGAA AAGATGCATA GAGGACTTTG AAGCCAGAGG
201 ACTGGGGGTG CTAAGAAGAG AGCTTGCATC TTTGACGCGT GATTTCATG
40 251  ATAAAGCAAA AGCAGAGACT TCTATGCTCA TTGAATGTCC TTGTATGGT
301  TTTTATTATA GTATTTCATCA GGAGGAACAA AGGCAAGGC AAGAAAGGCT
351  TCAAAAGATG GCTGAGCGCT ATAGGGACTG TAAACAAGTC TTGGAGGCTG
401  TCCAGGTGGA GCAAAAGAT ATGATATCTT CTAGAGTCGT TGTCGATGAC
451  AGCTACTTTG AAGAAGAAAA AGAAGAACAA AAGGTGGATA ACAGAAAGAA
45 501  AGAACAGGAC TAG

```

The PSORT algorithm predicts cytoplasm (0.3193).

The protein was expressed in *E.coli* and purified as a GST-fusion product (Figure 186A) and also in his-tagged form. The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (Figure 186B) and for FACS analysis.

These experiments show that cp6389 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 187

The following *C.pneumoniae* protein (PID 4376792) was expressed <SEQ ID 373; cp6792>:

```

5      1  VLQEHFFLSE DVITLAQQLL GHKLITTHEG LITSGYIVET EAYRGPDDKA
      51  CHAYNYRKTQ RNRAMYLKGG SAYLYRCYGM HLLNVTGTP EDIPHAVLIR
     101  AILPDQGKEL MIQRQWRDK PPHLLTNGPG KVCQALGISL ENNRQRLNTP
     151  ALYISKEKIS GTLTATARIG IDYAQYRDV PWRFLSPED SGKVLS*

```

The cp6792 nucleotide sequence <SEQ ID 374> is:

```

10      1  GTGTACAAG AACATTTTTT TCTATCGGAA GATGTAATTA CACTAGCGCA
      51  ACAGCTTTTA GGACATAAAC TCATCACAAC ACATGAGGGT CTGATAACTT
     101  CAGGTACAT TGTAGAAACC GAAGCGTATC GTGGCCCTGA TGACAAAGCA
     151  TGCCACGCCT ACAACTACAG AAAAATCAG AGGAACAGAG CGATGTACCT
     201  GAAAGGAGGC TCTGCTTACC TCTACCGTTG CTATGGCATG CATCACCTAT
15      251  TGAATGTTGT CACTGGACCT GAGGACATTC CCCATGCCGT CCTGATCCGG
     301  GCCATCCTTC CTGATCAAGG CAAAGAACTT ATGATCCAAC GCCGCCAATG
     351  GAGAGATAAA CCCCACACC TTCTACCAA TGGACCCGGA AAAGTGTGCC
     401  AAGCTCTAGG AATCTCTTTG GAAAACAATA GGCAACGCCT AAATACCCCA
     451  GCTCTCTATA TCAGCAAAGA AAAAATCTCT GGGACTCTAA CAGCAACTGC
20      501  CCGGATCGGC ATCGATTATG CTCAAGAGTA TCGTGATGTC CCATGGAGAT
     551  TTCTCTATC CCCAGAAGAT TCGGAAAAG TTTTATCTTA A

```

The PSORT algorithm predicts cytoplasm (0.180).

The protein was expressed in *E.coli* and purified as a his-tagged product (Figure 187A; lanes 2-4).

The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 187B) and for FACS analysis.

These experiments show that cp6792 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 188

The following *C.pneumoniae* protein (PID 4376868) was expressed <SEQ ID 375; cp6868>:

```

30      1  MVETVLHNFQ RYLSKYLYRV FRFPCRKKTF LSSHRVLARP SFPVDYCPGK
      51  IYDLQEIYEE LNAQLFQGA RLQIGWFRK ATRRGKSVVL GLFHENEQLI
     101  RIHRSLDRQE IPRPFMEYLV YHEMVHSVVP REYSLSGRSI FHGKKFKEYE
     151  QRFPLYDRAV AWEKANAYLL RGYKKRVGGG YGRA*

```

The cp6868 nucleotide sequence <SEQ ID 376> is:

```

35      1  ATGGTTGAAA CAGTACTTCA TAATTTCCAA CGTTATCTGA GCAAGTATCT
      51  CTATAGGGTA TTTCGCTTCC CATGTCGTAA AAAGACGTTT CTATCTTCGC
     101  ACAGGGTCTT TGCTCGTCCT TCATTTCCAG TAGACTACTG TCCGGGAAAG
     151  ATCTATGATT TGCAGGAGAT CTATGAGGAA TTGAATGCGC AGTTATTTCA
     201  AGGTGCACTG CGTTTACAGA TTGGTTGGTT CGGAAGGAAA GCTACCAGAA
40      251  AAGGCAAGAG TGTGTCTTTC GGATTGTTTC ATGAAAATGA ACAGTTAATT
     301  CGAATTCATC GTTCTTTAGA TCGGCAGGAA ATCCCAAGAT TTTTATGGA
     351  ATATCTTGTG TATCATGAAA TGGTTCATAG TGTAATCCCT AGAGAGTATT
     401  CTCTATCGGG GCGTTCGATT TTTCATGGTA AAAAGTTTAA AGAATACGAA
     451  CAACGTTTCC CTTGTATGA TCGTGCTGTT GCTTGGGAAA AGGCAAACGC
45      501  TTATTATTATG CGAGGGTATA AAAAAAGAGT AGGTGGAGGA TATGGCAGGG
     551  CATAG

```

The PSORT algorithm predicts bacterial cytoplasm (0.325).



The protein was expressed in *E.coli* and purified as a his-tag product (Figure 188A; lanes 2-3). The recombinant protein was used to immunise mice, whose sera were used in a Western blot (Figure 188B) and for FACS analysis.

These experiments show that cp6868 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 189

The following *C.pneumoniae* protein (PID 4376894) was expressed <SEQ ID 377; cp6894>:

```

1  MYKRCVLDKI LKGIVAGSLI LLYWSSDLLE RDIKSIKGNV RDIQEDIREI
51  SRVVKQQQTS QAIIPAAPGVM LAPKLVRDEA FALLFGDPSY PNLLSLDPYK
101 QOTLPELLGT NFHPHGILRT AHVGKPENLS PFNGFDYVVG FYDLCIPSLA
151 SPHVKGKYEFP SPDLAVKIEE HLVEDGSGDK EFHIYLRPNV FWRPIDPKAL
201 PKHVQLDEVF QRPHPVTAHD IKFFYDAVMN PYVATMRAVA LRSCYEDVVS
251 VSVENDLKL VVRWKAHTVIN REGKEERKVL YSAFNTLSL QPLPRFVYQY
301 FANGEKIIED ENIDTYRTNS IWAQNFTMHW ANNYIVSCGA YYFAGMDDEK
15 351 IVFSRNPDFY DPLAALIDKR FVYFKESTDS LFQDFKTGKI DISYLPNQR
401 DNFYSFMKSS AYNKQVAKGG AVRETVSADR AYTYIGWNCF SLFFQSRQVR
451 CAMNMAIDRE RIIEQCLDGQ GYTISGPFAS SSPSYNKQIE GWHYSPERAA
501 RLLEEGWID TDGDGIREKV IDGVIVPFRF RLCYVKSVT AHTIADYVAT
551 ACKBIGIECS LLGLDMADLS QAFDEKNFDA LLMGWCLGIP PEDPRALWHS
20 601 EGAMEKGSAN VVGFIHNEAD KIIDRLSYEY DLKERNRLYH RFHEITHREA
651 PYAFLEFSRHC SLLYKDYVKN IFVPHTRTDL IPEAQDETVN VTMVWLEKKE
701 DPCLSTS*

```

The cp6894 nucleotide sequence <SEQ ID 378> is:

```

1  ATGTATAAAA GATGTGTGCT AGATAAAATT TTAAAGGGGA TTGTCCGCCG
25 51  TTCTTTAATT TTGTTATACT GGTCTCAGCA CCTACTTGAA AGAGACATTA
101 AGTCGATAAA AGGTAAACGTA AGAGATATTC AAGAAGACAT TCGTGAAATC
151 TCACGCGTAG TGAAACAACA GCAGACATCA CAAGCTATCC CTGCGGCACC
201 TGGGGTGTAG CTCGCTCCTA AGCTCGTCAG AGACGAAGCT TTTGCTCTAC
251 TCTTTGGAGA TCCTAGTTAT CCTAATTTAC TTTCCTAGA CCCCTATAAA
30 301 CAGCAGACTC TTCCTGAAC TCTAGGAACA AATTTCCACC CTCATGGTAT
351 CCTACGCACT GCCCATGTCG GAAACCCCGA AAATCTGAGC CCTTTTAATG
401 GCTTTGATTA TGTCTGGGCG TTTTACGATC TCTGTATTCC TAGTTTAGCT
451 TCTCCCCACG TAGGGAATA CGAAGAATT TCTCCAGATC TCGCTGTGAA
501 AATAGAAGAA CATCTTGTG AAGATGGTTC TGGGGATAAA GAGTTTCACA
35 551 TCTATCTGAG GCCGAATGTT TTTTGGCGTC CTATAGATCC TAAGGCCCTT
601 CCAAAACACG TTCAGTTAGA CGAAGTATTT CAACGTCCTC ATCCTGTGAC
651 AGCTCATGAT ATTAAGTTT TCTACGACGC TGTATGAAC CCTTATGTAG
701 CAACCATGCG AGCAGTGGCT CTGCGCTCTT GTTATGAAGA TGTGGTTTCT
751 GTCTCAGTAG AAAACGATTT AAAATTAGTA GTCAGATGGA AAGCACACAC
40 801 GGTAATCAAT GAAGAAGGAA AGGAAGAGCG CAAAGTGCTC TACTCTGCAT
851 TTTCTAATAC CTTAAGCTTG CAGCCCCTCC CTAGATTTGT ATATCAGTAT
901 TTTGCTAACG GGGAAAAAAT CATTGAAGAT GAGAATATCG ATACCTACCG
951 AACCAATTCC ATTTGGGCGC AAAACTTCAC TATGCATTGG GCAAACAAC T
1001 ATATTTGTAAG TTGTGTGAGCC TACTACTTTG CAGGGATGGA TGATGAGAAA
45 1051 ATCGTGTTTT CTAGAAATCC TGACTTCTAT GATCCTCTTG CGGCTCTTAT
1101 TGACAAGCGT TTCGTCTATT TTAAGGAAAG CACAGACTCC CTATTCCAAG
1151 ATTTTAAGAC AGGGAAAATA GACATCTCTT ACCTTCCACC CAACCAAAGA
1201 GATAATTTCT ATAGTTTAT GAAAAGCTCC GCTTATAACA AACAGGTAGC
1251 TAAGGAGGGA GCCGTCGGTG AAACAGTCTC AGCAGATCGA GCATATACGT
50 1301 ACATAGGATG GAATTGCTTT TCATTATTTT TCCAAAGCCG ACAGGTGCGC
1351 TGTGCTATGA ACATGGCAAT CGATAGAGAG AGGATATATCG AACAGTGCTT
1401 GGATGGCCAA GGCTATACGA TTAGTGGGCC TTTTGCTTCG AGTTCTCCTT
1451 CTTATAATAA ACAGATCGAA GGGTGGCATT ATTCTCCAGA AGAAGCAGCT
1501 CGTCTCCTGG AAGAAGAGGG ATGGATAGAT ACCGATGGCG ATGGAATCCG
55 1551 AGAAAAAGTT ATCGATGGTG TGATTGTCCC GTTCCGTTTC CGTTTATGCT
1601 ATTATGTAAA GAGTGTACAC GCTCATACCA TTGCAGATTA CGTAGCTACT
1651 GCTTGTAAAG AAATCGGAAT CGAGTGTAGC CTTCTAGGAC TAGATATGGC
1701 CGATCTTTCG CAAGCTTTTG ATGAAAGAA TTTTCGATGCT CTTTTAATGG
1751 GATGGTGTTC AGGAATTCCT CCTGAGGATC CTAGGCTTTT ATGGCATTC T

```

-187-

```

1801 GAAGGGGCTA TGGAAAAGGG TTCAGCGAAT GTTGTAGGTT TCCATAATGA
1851 AGAAGCTGAT AAAATCATAG ACAGACTCAG CTACGAATAC GATCTGAAAG
1901 AACGTAATCG CCTGTACCAC CGTTTCCATG AAATTATTCA TGAGGAAGCT
1951 CCTTATGCTT TCTTGTTC TCACATTGT TCCTTACTTT ATAAGGATTA
5 2001 TGTAAGAAAT ATTTTCGTAC CTACACATAG AACAGATTTA ATTCCTGAAG
2051 CTCAGGATGA GACTGTCAAC GTAACATATG TATGGCTTGA GAAGAAGGAG
2101 GATCCGTGCT TAAGTACATC CTAA

```

The PSORT algorithm predicts inner membrane (0.162).

The protein was expressed in *E.coli* and purified as a his-tag product (Figure 189A) and also in GST/his form. The recombinant proteins were used to immunise mice, whose sera were used in a Western blot (Figure 189B) and for FACS analysis.

These experiments show that cp6894 is a surface-exposed and immunoaccessible protein, and that it is a useful immunogen. These properties are not evident from the sequence alone.

### Example 190

The following *C.pneumoniae* protein (PID 4377193) was identified in the 2D-PAGE experiment <SEQ ID 379; cp7193>:

```

1  MKRVITYKTIF CGLTLLTSL SCSLDPKGYN LETKNSRDLN QESVILKENR
51  ETPSLVKRLS RRSRLRFARR DQTQKDTLQV QANFKTYAEK ISEQDERDLS
101 FVVSSAAEKS SISLALSQGE IKDALYRIRE VHPLALIEAL AENPALIEGM
20 151 KMQGRDWIW NLFLTQLSEV FSQAWSQGI SEEDIAAFAS TLGLDSGTVA
201 SIVQGERWPE LVDIVIT*

```

A predicted leader peptide is underlined.

The cp7193 nucleotide sequence <SEQ ID 380> is:

```

1  ATGAAAAGAG TCATTATAA AACCATATTT TCGGGTTAA CTTTACTTAC
25 51  AAGTTTGAGT AGTTGTTCCC TGGATCCTAA AGGATATAAC CTAGAGACAA
101 AAAACTCGAG GGACTTAAAT CAAGAGTCTG TTATACTGAA GGAAAACCGT
151 GAAACACCTT CTCTTGTTAA GAGACTCTCT CGTCGTTCTC GAAGACTCTT
201 CGCTCGACGT GATCAAACTC AGAAGGATAC GCTGCAAGTG CAAGCTAACT
30 251 TTAAGACCTA CGCAGAAAAG ATTTTCAGAGC AGGACGAAAG AGACCTTTCT
301 TTCGTGTGCT CGTCTGCTGC AGAAAAGTCT TCAATTTCGT TAGCTTTGTC
351 TCAGGGTGAA ATTAAGGATG CTTTGTACCG TATCCGAGAA GTCCACCCCTC
401 TAGCTTTAAT AGAAGCTCTT GCTGAAAACC CTGCCTTGAT AGAAGGGATG
451 AAAAAGATGC AAGGCCGTGA TTGGATTGG AATCTTTTCT TAACACAATT
501 AAGTGAAGTA TTTTCTCAAG CTTGGTCTCA AGGGTTATC TCTGAAGAAG
35 551 ATATCGCCGC ATTTGCCTCC ACCTTAGGTT TGGACTCCGG GACCGTTGCG
601 TCCATTGTCC AAGGGGAAAG GTGGCCCGAG CTTGTGGATA TAGTGATAAC
651 TTAA

```

The PSORT algorithm predicts periplasmic (0.925).

This shows that cp7193 is an immunoaccessible protein in the EB and that it is a useful immunogen.

These properties are not evident from the protein's sequence alone.

It will be appreciated that the invention has been described by way of example only and that modifications may be made whilst remaining within the spirit and scope of the invention.

TABLE II – sequences of the primers used to amplify Cpn genes.

Orf ID	N-terminus final primer	C-terminus final primer
CP0014P	GCGTC CCG GGT CATATG AAGTCTCTCTTCCCA	GCGT CTC GAG ATGAAAGAGTTTTCGCG
CP0015P	GCGTCCCGGGTCATATG TCAGCTCTGTTTTCTGA	GCGT CTC GAG GAATTTGGTATTTTGCTC
CP0016P	GCGTCCCGGGTCATATG GCCGATCTCACATTAG	GCGT CTC GAG GTCCAAGTTAAGGTAGCA
CP0017P	GCGT CCG GGT CATATG GGTATCAAGGAACTG	GCGT CTC GAG AAATCCGAATCTTCC
CP0019P	GCGTCCCGGGTCAT ATGCAAGACTCTCAAGACTATAG	GCGT CTC GAG AAATCCGTATTTTACCC
CP6260P	GCGTC CCG GGT GCTAGCACTACGATTCTTTAACCC	GCGT CTC GAG AAAACGAAATTTGCTTC
CP6397P	GCGTC CCG GGT CATATGTTTAACTGCTAAAAATCTATT	GCGT CTC GAG ATGAAAGAAGAGTCTCTCG
CP6456P	GCGTC CCG GGT CATATG TCATCTCTGTAAATAACA	GCGT CTC GAG CTGACCATCTCTCTGT
CP6466P	GCGTC CCG GGT CAT ATG TGCAAGGAGTCCAGT	GCGT CTC GAG ATTTTCTCTAGCATAAAG
CP6467P	GCGTC CCG GGT CAT ATG TGTTCCCATCCCAA	GCGT CTC GAG TAGTTTTCTATATAACGAAAGTCT
CP6468P	GCGTC CCG GGT CAT ATG TGCTCTCTCTACTCTTC	GCGT CTC GAG GGGGAAATAGGTATATTTGA
CP6489P	GCGTC CCG GGT CAT ATG AGCTGCTCAAGCAA	GCGT CTC GAG ACTTAAGATATCGATATTTTGA
CP6552P	GCGTC CCG GGT CAT ATG TGCCATAAGGAAGATG	GCGT CTC GAG ACCATTGTCTTGAGTCAT
CP6567P	GCGTC CCG GGT CAT ATG ACCTCACCGATCCCC	GCGT CTC GAG AGAAGCCGGTAGAGGC
CP6576P	GCGTC CCG GGT CAT ATG ACTGAAAAAGTTAAAGAAGG	GCGT CTC GAG GAA CATGCCCCCTAA
CP6727P	GCGTC CCG GGT CATATGCTACATCCACTAATGGC	GCGT CTC GAG GAAAGAATAACGAGTTCC
CP6729P	GCGTC CCG GGT CAT ATGGCAGATGCTTCTTTATC	GCGT CTC GAG GAATGAATATCTTAGCC
CP6731P	GCGTC CCG GGT CATATGGCTGTGTGTGAATCAAT	GCGTC CAT GGC GGC CGC GAACTGGAACTTACCTCC
CP6736P	GCGTC CCG GGT GCT AGCGTAGAAGTTATCATGCTTT	GCGTC CAT GGC GGC CGC AAATCGTAATTTGCTTC
CP6737P	GCGT GGA TCC CAT ATG GAGACTAGACTCGGAGG	GCGT CTC GAG AAATGTGATTTTAGTCC
CP6751P	GCGTC CCG GGT GCT AGC AATGAAGGTCTCCAAT	GCGT CTC GAG AAATCTCATTTCTACTCGC
CP6752P	GCGTGA ATT CAT ATGTTCCGGATGACPTCT	GCGT CTC GAG GAATTTTAAGGTACTTCTCTG
CP6753P	GCGTC CCG GGT GCT AGCACTCCCTACTCTCATAGAG	GCGT CTC GAG AAACCTTAAAGTCTGCTC
CP6767P	GCGTC CCG GGT CAT ATG ATAAAACAATAGGCCGT	GCGT CTC GAG TTCGTAAGCACTTTCAGA
CP6828P	GCGTC CCG GGT CAT ATG AAGCAGATGCGCTTTT	GCGTC CAT GGC GGC CGC GAAACTAAGGAGAGGC
CP6830P	GCGTC CCG GGT CAT ATG GATCCCGCTCTGTT	GCGTC CAT GGC GGC CGC GAATACAAACGGATCC
CP6832P	GCGTC CCG GGT CAT ATG CATAAAGTAATAGTTTTCATT	GCGT CTC GAG TAAACTAGAAAAAGTCTGC
CP6848P	GCGTC CCG GGT CAT ATG TCATCAAATCTACATCCC	GCGT CTC GAG AACCGAGCTATTTTAC
CP6849P	GCGTC CCG GGT GCT AGC AGCGGGGGTATAGAG	GCGT CTC GAG ATACACGTGGGTATTTTC
CP6850P	GCGTC CCG GGT CAT ATG TGCCGCAATGTAGAT	GCGT CTC GAG CTGTTTGCATCTGCC
CP6854P	GCGTC CCG GGT GCT AGC TCAATAGCTATTGCAAG	GCGT CTC GAG TTATCGAAATGTCTTTG
CP6879P	GCGTC CCG GGT CAT ATG GCAACACCCGCTCAA	GCGTC CAT GGC GGC CGC TCCTTGAATTTGCTCTTGC
CP6894P	GCGTC CCG GGT CAT ATG TATAAAGATGTGTCTAGA	GCGT CTC GAG GGATGTACTTAAGCAGC
CP6900P	GCGTC CCG GGT CAT ATG AAGATAAAATTTCTTGGAAAG	GCGT AAG CTT GGGAGACGATACCG
CP6952P	GCGTC CCG GGT CAT ATG CTCTCGGATCAATATATAGG	GCGT CTC GAG TCGAATTTCTTTTTFAGC
CP7034P	GCGTC CCG GGT CAT ATG AAAAACAGGTATATCAATG	GCGT AAG CTT AAACGCTGAAATTTATACC
CP7090P	GCGTC CCG GGT CAT ATG TGTAGCCTTTCCCTT	GCGT CTC GAG GCGTGCATGAATCTTA
CP7091P	GCGTC CCG GGT CAT ATG GAAGAATTAGAAGTTGTGT	GCGT CTC GAG TAGTGTCTCTTTATCGGT
CP7170P	GCGTC CCG GGT CAT ATG CTAGGGGCTGGAAACC	GCGT AAG CTT AAATGCGAGACCTGACG
CP7228P	GCGTC CCG GGT CAT ATG ACTGCTGTTCTTATCTTACA	GCGT CTC GAG ATCTGAAAGCGGAGG
CP7249P	GCGTC CCG GGT CAT ATG ATCCCATCCCTACC	GCGT CTC GAG ATCAGGTTGCTGAGACTT
CP7250P	GCGTC CCG GGT CAT ATG AATCTTTCAAACAGGTCT	GCGT CTC GAG ATTTTCTTAGAGAGACTCTC
CP0018P	GTGCGT CATATG GCAACCACTCCACTAA	ACTCGCTA GCGGCCGC TAATGAGGTCCCCAG
CP6270P	GTGCGT CATATG AATTATATAGGAGCTGCT	ACTCGCTA GCGGCCGC AAATTTGATTTTGCTACC
CP6735P	GTGCGT CATATG GCAGCACAAAGTTGTATAT	ACTCGCTA GCGGCCGC TGGCGTAGAAGTGATC
CP6898P	GTGCGT CATATG TTGCCTGTAGGGAAAC	ACTCGCTA GCGGCCGC GAATCTGAACTGACCAGA
CP7033P	GTGCGT CATATG GTTAATFCTATTTGGTCCA	ACTCGCTA GCGGCCGC TTGGAGATAACCAATATA
CP7287P	GTGCGT CATATG TTACACAGCTCAGAACTAGA	ACTCGCTA GCGGCCGC GAAATAATACGGATACCA
CP0010P	GTGCGT CATATG GCAACTGCTGAAAATATA	GCGT CTCGAG GAATTTGAACTTACCC
CP0468P	GTGCGT GCTAGC ATTTTATATGACAACTCTAT	GCGT CTCGAG AAATGTGCAATGACTCT
CP6272P	GTGCGT CATATG TTGACTCATCAAGAGGCT	GCGT CTCGAG GAAGGGAGGTTTTFATGGT
CP6273P	GTGCGT CATATG ACATATCTGGAAGCTC	ACTCGCTA GCGGCCGC CTCCACAATTTTATG
CP6362P	GTGCGT CATATG CCCTTTGATATTAATTTATATA	GCGT CTCGAG TCGTTTCCAAATCCA
CP6372P	GTGCGT CATATG AAACAACACTATTTCTAAATA	GCGT CTCGAG TTCTTTGTTGTTTCT
CP6390P	GTGCGT CATATG CGAGAGGTGCTTAAG	ACTCGCTA GCGGCCGC TCTCTAGACAGCCTT
CP6402P	GTGCGT CATATG AATGTTGCGGATCTCCCTT	GCGT CTCGAG GAAGGGGTTGCGCGT
CP6446P	GTGCGT CATATG TGTAAATCAAAAGCCCTCTT	GCGT CTCGAG GGGCTGAGGAGGAAC
CP6520P	GTGCGT GCTAGC AAACACTACCTATCATTTTCT	GCGT CTCGAG CAGAAAGGCTTTTCTT
CP6577P	GTGCGT CATATG AATTAGGCTATGTTAATTTA	GCGT CTCGAG GTTTTGTGTTTGTGAAGA
CP6602P	GTGCGT CATATG GCAGCATCAGGAGGCA	GCGT CTCGAG TGACCAAGGATAGGGTTTAG

CP6807P	GTGCGT CATATG CCTCGTGGTGACACTTT	GCGT CTCGAG CGCTGCTTCTTGCTC
CP6815P	GTGCGT CATATG TGCTCTCAAAAAACGACAA	GCGT CTCGAG TGAAGAGGCGCCATC
CP6824P	GTGCGT CATATG GATGCGAAAATGGGA	GCGT CTCGAG TCTTTGACATTCAAGAGC
CP6872P	GTGCGT CATATG ATTCTACCATGTTAATG	GCGT CTCGAG GTCATACAATTTCTTATATA
CP6879P	GTGCGT CATATG TGCACTCACTTAGGCT	GCGT CTCGAG CGAGTAGTTAGCAGAAAC
CP6717P	GTGCGT GCTAGC AAGACAATCGTAGCTTCA	ACTCGCTA GCGGCCGC GGCTGGCATATAGGT
CP6784P	GTGCGT GCTAGC AAATCAAGATGTTCTATTGATA	GCGT CTCGAG TCCAAAACAAACCTCT
CP6802P	GTGCGT CATATG TCGTAAGTTATATTATTCCTT	GCGT CTCGAG CAGTCGGGCTTGTTG
CP6847P	GTGCGT CATATG TCGGATCTTTTACGAG	GCGT CTCGAG TTTTCTACACTGTGTATAATAA
CP6884P	GTGCGT CATATG AATCAGCTGCTTCT	GCGT CTCGAG AGAGAAGGTAATGTACC
CP6886P	GTGCGT CATATG TGTCTACTTATTATCTATCTCTAC	GCGT CTCGAG TTCAGAAAAATGGCT
CP6890P	GTGCGT CATATG TCCCCACGACGACAA	GCGT CTCGAG TCTCGACGATTTAGC
CP6860P	GTGCGT CATATG TGTGACGTACGGTCTA	ACTCGCTA GCGGCCGC TTCACCTTGATTTCCT
CP6868P	GTGCGT CATATG TCGGATGCAAAAC	ACTCGCTA GCGGCCGC GGAAGTATGCTTAGATATT
CP6869P	GTGCGT CATATG TGCTGTGGTTACTCTATT	ACTCGCTA GCGGCCGC AAAAAGGTCATAGTATACCT
CP7005P	GTGCGT CATATG AAAACTGTGATATTGAACA	GCGT CTCGAG CTGAGCTTCTATTTCATTAT
CP7072P	GTGCGT CATATG CCCATTTATGGGAAA	GCGT CTCGAG GTTGAGCAAAGGTTTG
CP7101P	GTGCGT CATATG TATTCGTGTACAGCAA	GCGT CTCGAG GAAAAATTCCTTAGGGAG
CP7102P	GTGCGT CATATG GCGCTAAAGCAAAT	GCGT CTCGAG TGAAAAATGAAGGATGGT
CP7105P	GTGCGT GCTAGC AGTCTATATCAAAAATGGTG	GCGT CTCGAG ATCTTTTATTTGGTTATCT
CP7106P	GTGCGT CATATG AAAGATTTGGGGACTCT	GCGT CTCGAG GAATCCTAAGGCATACCTA
CP7107P	GTGCGT GCTAGC AGTATAGTCAGAAATTCGCA	GCGT CTCGAG GAAGCTAAGATTATAGCTACTTT
CP7108P	GTGCGT GCTAGC GCGGCCCTTTCCA	ACTCGCTA GCGGCCGC TTTATGTATATGGAACAGATAGG
CP7109P	GTGCGT CATATG GGACATTTTATTTGATATTG	ACTCGCTA GCGGCCGC ATCATCAAGGTAGATAAAG
CP7110P	GTGCGT CATATG GGTATTTGCTATGTAAATTACA	GCGT CTCGAG TTCTGATTTGGACTCCA
CP7127P	GTGCGT CATATG GTGGCTTTAACGATAGC	ACTCGCTA GCGGCCGC GCAGCCATCGTATTC
CP7130P	GTGCGT CATATG TTCAATATGCGAGG	GCGT CTCGAG CTCTTTATTGAACTTTG
CP7140P	GTGCGT CATATG ACAGCCGGAGCAGCT	GCGT CTCGAG AGCACCCCTCAATTTCAATG
CP7182P	GTGCGT CATATG GGATATGTTTCTATGTGATC	GCGT CTCGAG GCTACTAAATCGAATCGA
CP6262P	GTGCGT CATATG ATCCCTGGATTAAAGTTCA	ACTCGCTA GCGGCCGC TTCCTGGGAGCTTGA
CP6266P	GTGCGT CATATG TACCAGGAGAACTAAGAT	ACTCGCTA GCGGCCGC GATTTCTTCTTCAGCTC
CP6296P	GTGCGT CATATG GAGGAGGTGCTGAGTAT	ACTCGCTA GCGGCCGC ATGTTCTTTTACTCTTTCT
CP6419P	GTGCGT CATATG GCTCCAGTCCGTGTT	GCGT CTCGAG AAGTGTTCGTGGAAGT
CP6601P	GTGCGT CATATG AATAAGCTACTCAATTTCTGT	GCGT CTCGAG GAAAAATCTGAATTTCTCT
CP6639P	GTGCGT CATATG TTAATAATCAAGCAATCA	GCGT CTCGAG AGGAACTAAAACTCATCT
CP6664P	GTGCGT GCTAGC GTTTTATTTTCACTGCTCAA	ACTCGCTA GCGGCCGC CTTAGAAAGACTATTTCTAAGTA
CP6696P	GTGCGT CATATG TGCCTGATAAATGGG	GCGT CTCGAG ATTCACTCTGTAAGAAAT
CP6757P	GTGCGT CATATG GCAGTTGGTGGCGT	ACTCGCTA GCGGCCGC CTGTCCCTCTGGAGC
CP6790P	GTGCGT GCTAGC AGTGAACACAAAAATCA	ACTCGCTA GCGGCCGC CTATCTGCTGTTATCAATA
CP6814P	GTGCGT CATATG CATGACGACTTCTAAG	GCGT CTCGAG TACAGCTGCGCGA
CP6834P	GTGCGT CATATG GTTATGGGAACCTATATCG	GCGT CTCGAG TACATTGTATTTGATTTCAG
CP6878P	GTGCGT CATATG AACGTCCCTGATTC	GCGT CTCGAG GCTAGCGGCTCTTC
CP6892P	GTGCGT CATATG CAGAAGCATCCTTCCT	ACTCGCTA GCGGCCGC TCCTCTTTAGGAAATGG
CP6909P	GTGCGT CATATG TCTCTTTAGGAAATGG	GCGT CTCGAG CAGTGCCAAGTAGGGA
CP7015P	GTGCGT CATATG GCAGTACGATTAATTTGTTG	GCGT CTCGAG TTTATTTGATGCTATTTTATATTTT
CP7035P	GTGCGT GCTAGC AGCAGAAAAAGACAATGA	GCGT CTCGAG ATTTTGAGTGTCTTGCA
CP7073P	GTGCGT CATATG ATTACCAATAATCACGTG	GCGT CTCGAG TATCCATCGACTTATAGC
CP7085P	GTGCGT GCTAGC TGTATTTTCCCTTACGTA	ACTCGCTA GCGGCCGC GGATTTCTGCATACTCTG
CP7092P	GTGCGT CATATG TCTCTCTTCTCTAAAAAA	GCGT CTCGAG GGATTCATTACTGACCA
CP7093P	GTGCGT CATATG AAATACCGCTTACG	GCGT CTCGAG ATTCTGTAGGCTACGT
CP7094P	GTGCGT CATATG GTACACTTCTCTCATAAACC	GCGT CTCGAG TAAGTTTGTATTGCGGTAT
CP7132P	GTGCGT CATATG TTGTTATTAGGACTTTAGGA	GCGT CTCGAG TTTCCCAACCGCA
CP7133P	GTGCGT CATATG GCTGCGAATGCTC	GCGT CTCGAG TAAATTAATACTCTTTGAAGG
CP7177P	GTGCGT CATATG CCTACTCAAGTTAAAACAGA	GCGT CTCGAG AAGTTTATATTTCAGCACTT
CP7184P	GTGCGT GCTAGC CATATAGGATTTTGCCA	GCGT CTCGAG GTACTTAGCAAAGCGAT
CP7206P	GTGCGT GCTAGC AAGAAGCTATATCACCTTA	GCGT CTCGAG CACACCGAGGAAAC
CP7222P	GTGCGT CATATG GTAGTTTCAGAGAAAAAGTC	GCGT CTCGAG ACGTATGCGCAACTG
CP7223P	GTGCGT CATATG GAAGTATTAGACCGCTCT	GCGT CTCGAG CGAGAAAAAGCTTCC
CP7224P	GTGCGT CATATG ATGAAGAAAAATTCGAAA	ACTCGCTA GCGGCCGC TAAGCATTCACAAATGA
CP7225P	GTGCGT CATATG CATATTTTGCTTGATGCT	GCGT CTCGAG TCTTTTAACTAAATCTTGTTCTT
CP7303P	GTGCGT CATATG CTGTCTATTTGTTTGATCC	GCGT CTCGAG AAAATATACGGAACCTCG
CP7304P	GTGCGT GCTAGC GAAGTTTATAGTTTTCCTC	GCGT CTCGAG TTTTGTATCTCTTAAGAAG
CP7305P	GTGCGT CATATG GAAGTTTATAGTTTTCACCTT	GCGT CTCGAG ACTCCTTGAGAGGGAA
CP7307P	GTGCGT CATATG CTAAATCATGCTAAAAAGC	ACTCGCTA GCGGCCGC CTCTTTTATTTTAGAAGCT

CP7342P	GTGCGT CATATG AAAAAAAAAATTATTTCTACT	ACTCGCTA GCGGCCGC CACACTCTGTTCTCTG
CP7347P	GTGCGT CATATG TTTTCTAAGGATTGACTAA	GCGT CTCGAG CGAAGCAGAAGTCGT
CP7353P	GTGCGT CATATG AATATGCGTGTTCCTTCT	GCGT CTCGAG GGGGCGTAGGTTGTA
CP7193P	GTGCGT CATATG TGTTCCTGGATCCT	ACTCGCTA GCGGCCGC AGTTATCACTATATCCACAAG
CP7248P	GTGCGT GCTAGC CTTGAACATTCTAAACAAGAT	GCGT CTCGAG ACGTAGTTTAAGAGCAGACT
CP7261P	GTGCGT CATATG TGTCTATCTGCCTACATAG	GCGT CTCGAG TTTTGATGCTTCTTTCA
CP7280P	GTGCGT CATATG GACCAGAAAATTGAAAA	GCGT CTCGAG AGAGGCTTCTGAGTGC
CP7302P	GTGCGT CATATG AATTTCATTGTAGTGTAGT	GCGT CTCGAG GAACAGTTCGATTGTG
CP7306P	GTGCGT CATATG CTTCCCTTATCAGGGCA	ACTCGCTA GCGGCCGC TTCTTCAGGTTTCAGG
CP7367P	GTGCGT GCTAGC CGTTATGCCGAGGTC	GCGT CTCGAG TTCGTGCATTGTG
CP7408P	GTGCGT CATATG TTGAAAATCCAGAAAAA	GCGT CTCGAG ATTCAATTTTCGGAAGAG
CP7409P	GTGCGT CATATG AGACGTTATCTTTTCATGGT	GCGT CTCGAG CCCTTGTCTCTTACATAG
CP8733P	GTGCGT ACTAGT TGTCACTACAGTCACTAG	GCGT CTCGAG GAATCGGAGTTTGTA
CP8728P	GTGCGT ACTAGT AAGTCCTCTGTCTCTTGG	GCGT CTCGAG GAAACAAAACCTTAGAGCCC

TABLE III – Proteins with best results in FACS analysis

cp number	Molecular Weight (kDa)		Fusion type
	Theoretical	Western Blot	
6260	97.5	94; 70	GST
6270	87.5	-	GST
6272	78.0	90	GST
6273	58.6	74; 64; 50	GST
6296	31.1	-	GST
6390	88.9	102	GST
6456	42.5	89; 67; 45	GST
6466	57.5	59; 56	His
6467	59.0	67	GST
6552	28.4	50; 27	GST
6576	86.0	79; 70; 62; 45	GST
6577	17.3	12	GST
6602	43.4	53; 42; 34	GST
6664	54.5	104; 45	GST
6696	47.9	95; 53	GST
6727	130.0-142.9	123; 61; 39	His
6729	94.8	multiple bands	GST
6731	95.5	97	GST
6733	97.1	104	His
6736	100.1	98; 93; 66; 60	GST
6737	101.2	multiple bands	GST
6751	100.2	95; 71	GST
6752	102.1	97; 48	His
6767	29.1	28	GST
6784	32.9	35	GST
6790	71.3	multiple bands	His
6802	29.7	-	GST
6814	29.6	28	GST

-191-

6830	177.4	174; 91; 13	GST
6849	57.3	multiple bands	GST
6850	7.4-9.4	61; 14; 8	GST
6854	42.2	-	GST
6878	40.4	-	GST
6900	28.0	-	GST
6960	25.6	75; 35	GST
6968	34.6	83; 53; 35	GST
6998	39.3	multiple bands	GST
7033	68.2	multiple bands	GST
7101	113	105	GST
7102	63.4	-	GST
7105	29.2	30	GST
7106	39.5	72;46	GST
7107	71.4	67; 31	His
7108	35.9	35	GST
7111	46.1	51	GST
7132	17.9	57; 47; 17	His
7140	36.2-29.8	50; 38; 34	GST
7170	34.4	77; 33	GST
7224	39.4	40	GST
7287	167.3	180	GST
7306	50.1	50	GST

TABLE IV – FACS-positive proteins not found in *C.trachomatis*

cp7105	cp6390
cp7106	cp6784
cp7107	cp6296
cp7108	

TABLE V – Proteins identified by MALDI-TOF following 2D electrophoresis

cp6270	cp6733	cp6900
cp6552	cp6736	cp6960
cp6576	cp6737	cp6998
cp6577	cp6752	cp7033
cp6602	cp6767	cp7108
cp6664	cp6784	cp7111
cp6727	cp6790	cp7170
cp6728	cp6830	cp7287
cp6729	cp6849	cp7306

**CLAIMS**

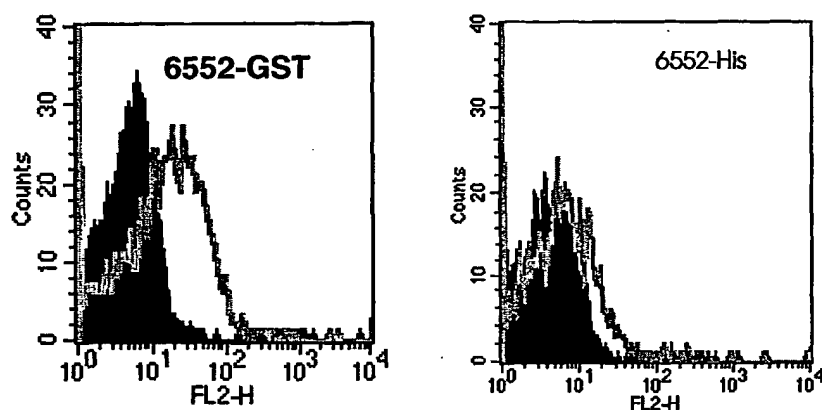
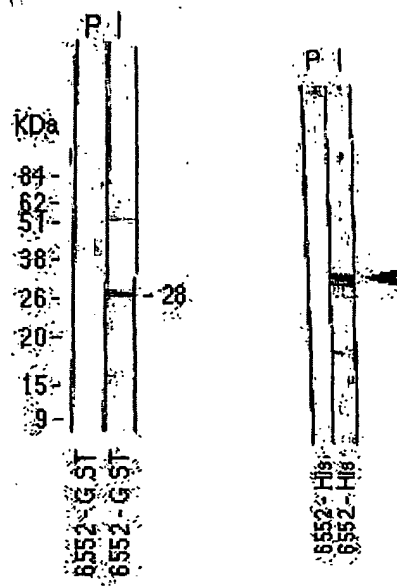
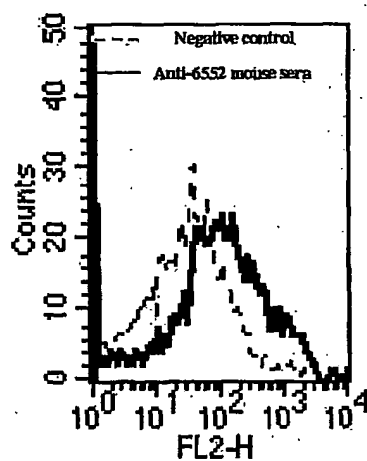
1. A protein comprising an amino acid sequence selected from the group consisting of SEQ IDs 97, 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95, 99, 101, 103, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127, 129, 131, 133, 135, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, 193, 195, 197, 199, 201, 203, 205, 207, 209, 211, 213, 215, 217, 219, 221, 223, 225, 227, 229, 231, 233, 235, 237, 239, 241, 243, 245, 247, 249, 251, 253, 255, 257, 259, 261, 263, 265, 267, 269, 271, 273, 275, 277, 279, 281, 283, 285, 287, 289, 291, 293, 295, 297, 299, 301, 303, 305, 307, 309, 311, 313, 315, 317, 319, 321, 323, 325, 327, 329, 331, 333, 335, 337, 339, 341, 343, 345, 347, 349, 351, 353, 355, 357, 359, 361, 363, 365, 367, 369, 371, 373, 375, & 377.
2. A protein having 50% or greater sequence identity to a protein according to claim 1.
3. A protein comprising a fragment of an amino acid sequence selected from the group consisting of SEQ IDs 97, 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95, 99, 101, 103, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127, 129, 131, 133, 135, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, 193, 195, 197, 199, 201, 203, 205, 207, 209, 211, 213, 215, 217, 219, 221, 223, 225, 227, 229, 231, 233, 235, 237, 239, 241, 243, 245, 247, 249, 251, 253, 255, 257, 259, 261, 263, 265, 267, 269, 271, 273, 275, 277, 279, 281, 283, 285, 287, 289, 291, 293, 295, 297, 299, 301, 303, 305, 307, 309, 311, 313, 315, 317, 319, 321, 323, 325, 327, 329, 331, 333, 335, 337, 339, 341, 343, 345, 347, 349, 351, 353, 355, 357, 359, 361, 363, 365, 367, 369, 371, 373, 375, & 377.
4. A nucleic acid molecule which encodes a protein according to any one of claims 1 to 3.
5. A nucleic acid molecule according to claim 4, comprising a nucleotide sequence selected from the group consisting of SEQ IDs 98, 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, 194, 196, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 228, 230, 232, 234, 236, 238, 240, 242, 244, 246, 248, 250, 252, 254, 256, 258, 260, 262, 264, 266, 268, 270, 272, 274, 276, 278, 280, 282, 284, 286, 288, 290, 292, 294, 296, 298, 300, 302, 304, 306, 308, 310, 312, 314, 316, 318,

320, 322, 324, 326, 328, 330, 332, 334, 336, 338, 340, 342, 344, 346, 348, 350, 352, 354, 356, 358, 360, 362, 364, 366, 368, 370, 372, 374, 376, & 378.

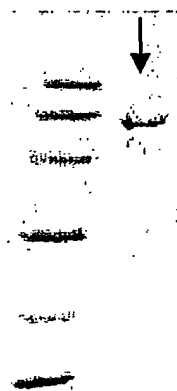
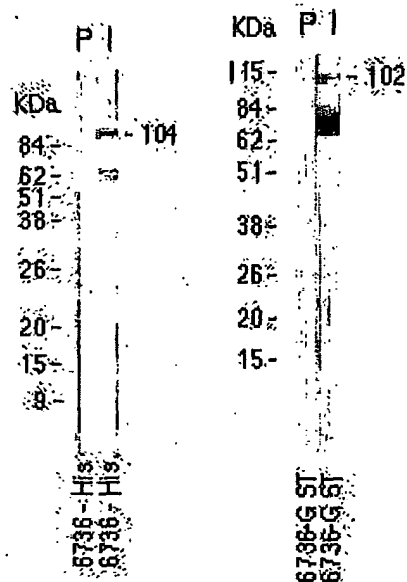
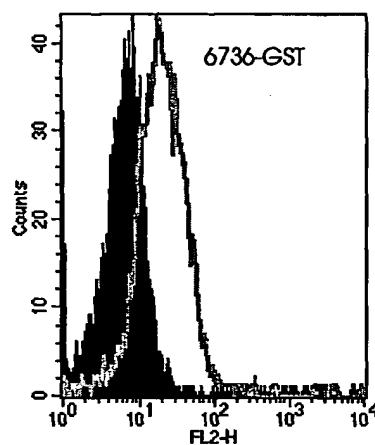
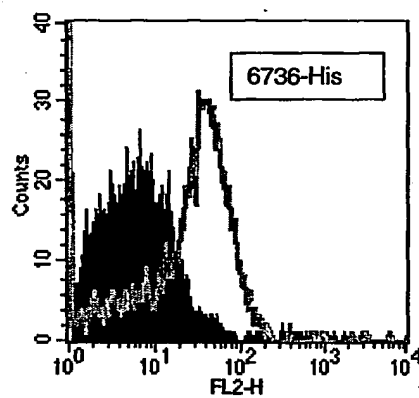
- 5 6. A nucleic acid molecule comprising a fragment of a nucleotide sequence selected from the group consisting of SEQ IDs 98, 2, 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, 194, 196, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, 228, 230, 232, 234, 236, 238, 240, 242, 244, 246, 248, 10 250, 252, 254, 256, 258, 260, 262, 264, 266, 268, 270, 272, 274, 276, 278, 280, 282, 284, 286, 288, 290, 292, 294, 296, 298, 300, 302, 304, 306, 308, 310, 312, 314, 316, 318, 320, 322, 324, 326, 328, 330, 332, 334, 336, 338, 340, 342, 344, 346, 348, 350, 352, 354, 356, 358, 360, 362, 364, 366, 368, 370, 372, 374, 376, & 378.
- 15 7. A nucleic acid molecule comprising a nucleotide sequence complementary to a nucleic acid molecule according to any one of claims 4 to 6.
8. A nucleic acid molecule comprising a nucleotide sequences having 50% or greater sequence identity to a nucleic acid molecule according to any one of claims 4 to 7.
9. A nucleic acid molecule which can hybridise to a nucleic acid molecule according to any one of claims 4 to 8 under high stringency conditions.
- 20 10. A composition comprising a protein or a nucleic acid molecule according to any preceding claim.
11. A composition according to claim 10 being a vaccine composition.
12. A composition according to claim 10 or claim 11 for use as a pharmaceutical.
- 25 13. The use of a composition according to claim 10 in the manufacture of a medicament for the treatment or prevention of infection due to *Chlamydia* bacteria, particularly *Chlamydia pneumoniae*.



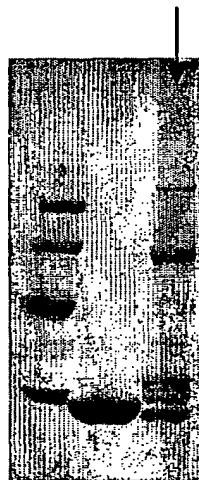
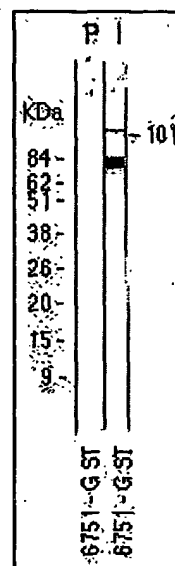
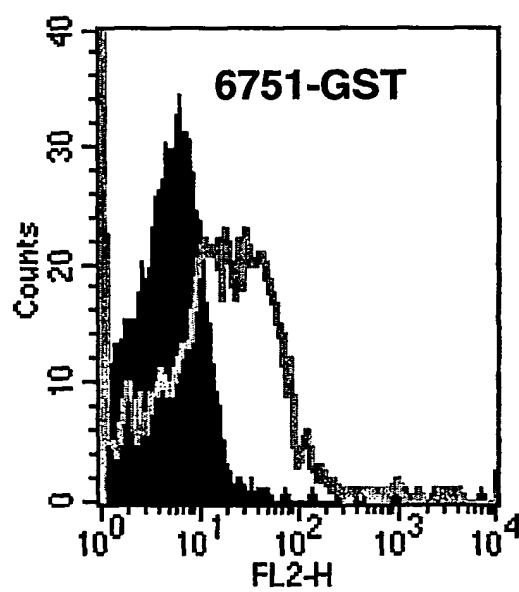
1/169

**FIGURE 1****Fig. 1A****Fig. 1B****Fig. 1C**

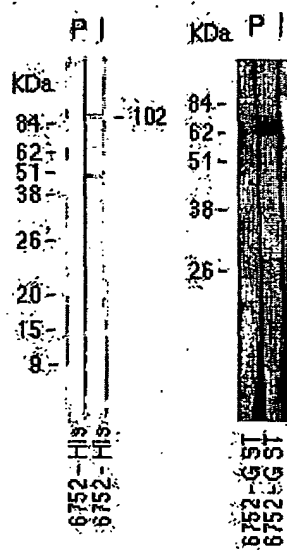
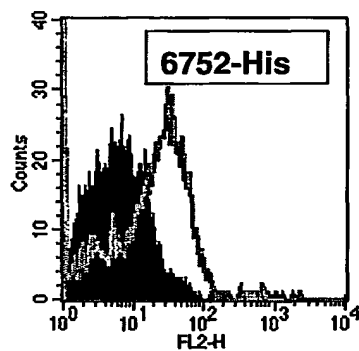
2/169

**FIGURE 2****Fig. 2A****Fig. 2B****Fig. 2C**

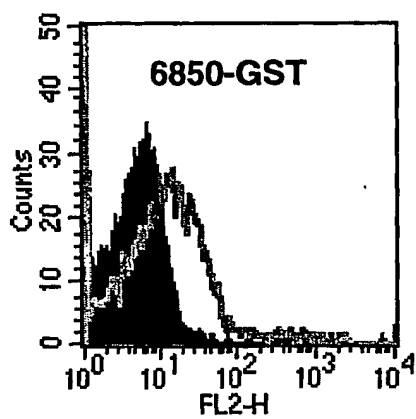
3/169

**FIGURE 3****FIG. 3A****FIG. 3B****FIG. 3C**

4/169

**FIGURE 4****FIG. 4A****FIG. 4B****FIG. 4C**

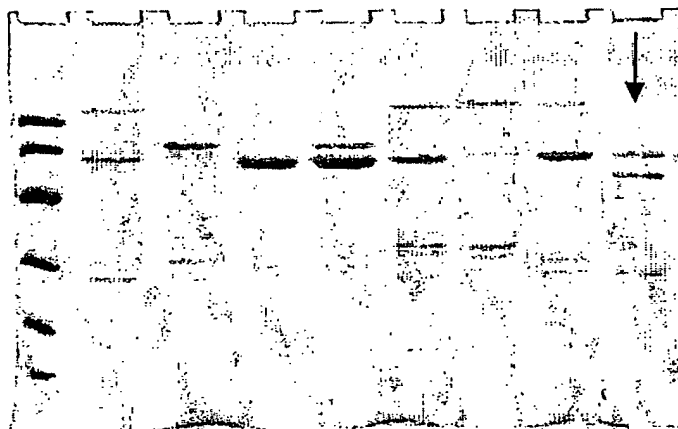
5/169

**FIGURE 5****Fig. 5A****Fig. 5B****Fig. 5C**

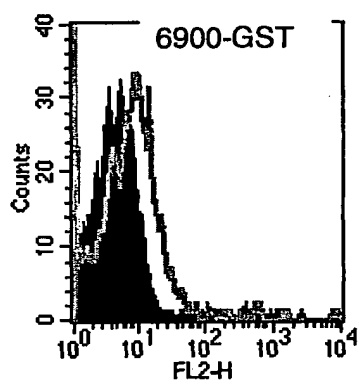
6/169

# **FIGURE 6**

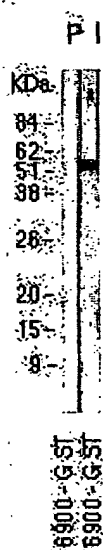
**FIG. 6A**



**FIG. 6B**

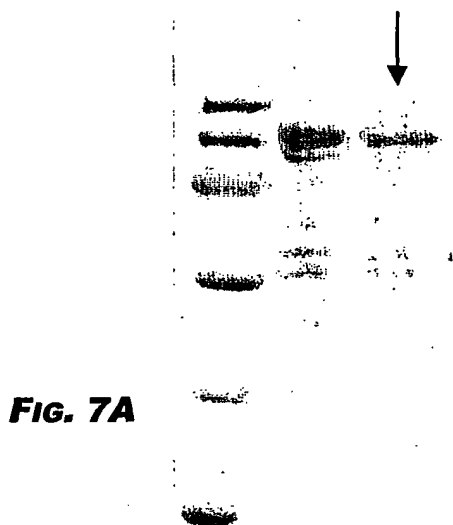


**FIG. 6C**

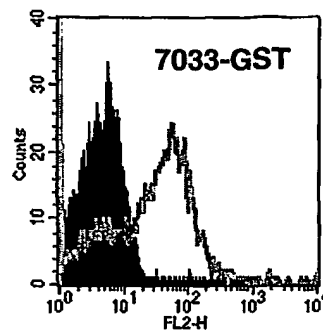


7/169

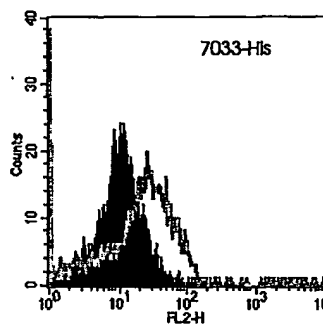
**FIGURE 7**



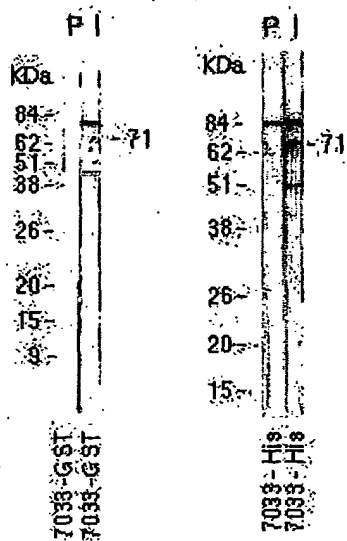
**FIG. 7A**



**FIG. 7B**



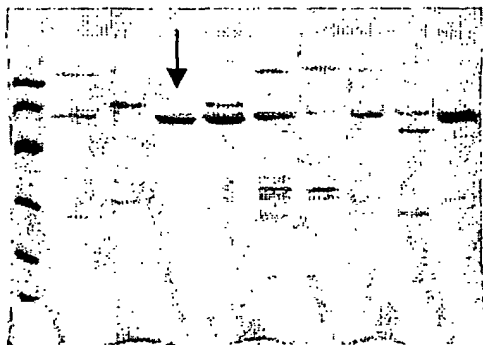
**FIG. 7c**



8/169

**FIGURE 8**

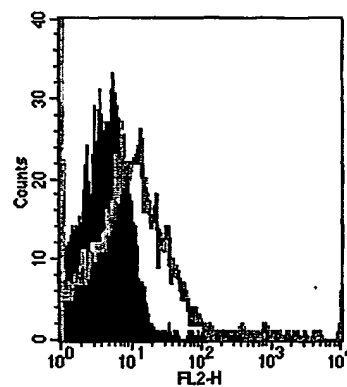
**Fig. 8A**



**Fig. 8B**



**Fig. 8C**

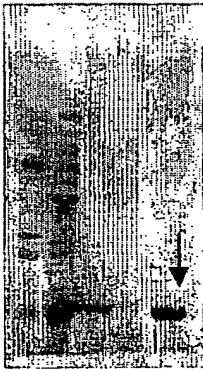




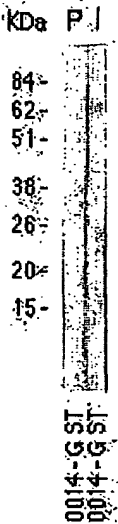
9/169

**FIGURE 9**

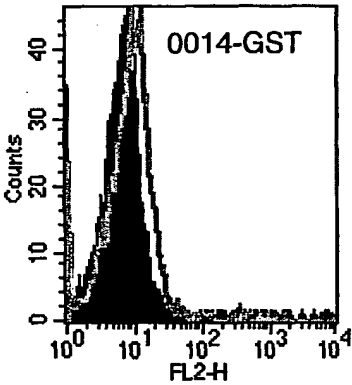
**FIG. 9A**



**FIG. 9B**



**FIG. 9C**



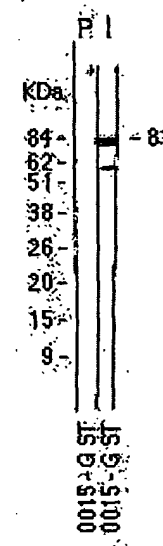
10/169

**FIGURE 10**

**FIG. 10A**



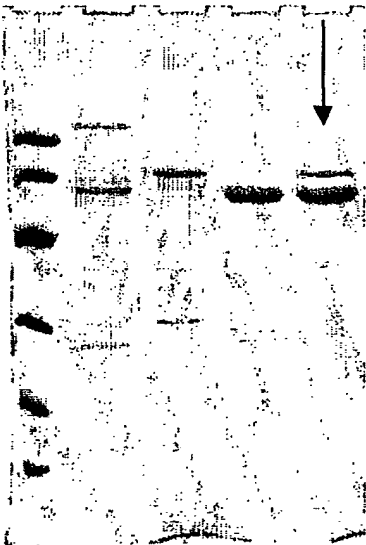
**FIG. 10B**



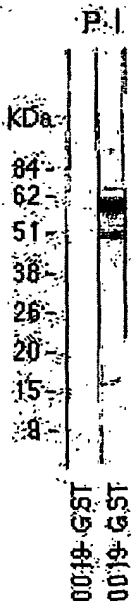
11/169

**FIGURE 11**

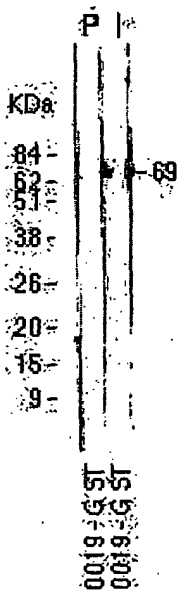
**Fig. 11A**



**Fig. 11B**



**Fig. 11C**



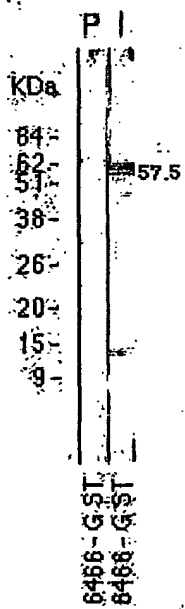
12/169

**FIGURE 12**

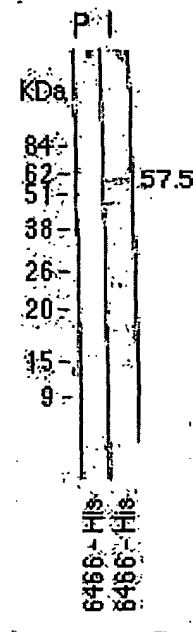
**FIG. 12A**



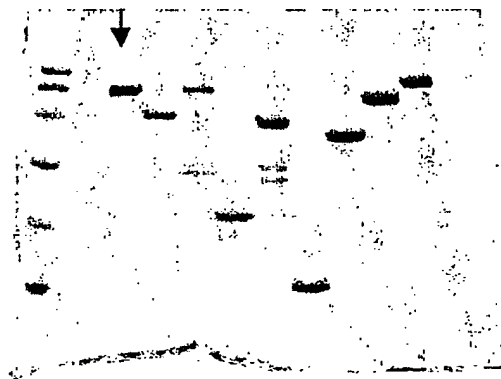
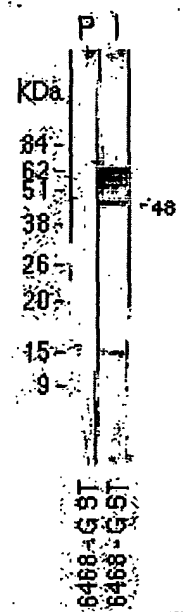
**FIG. 12B**



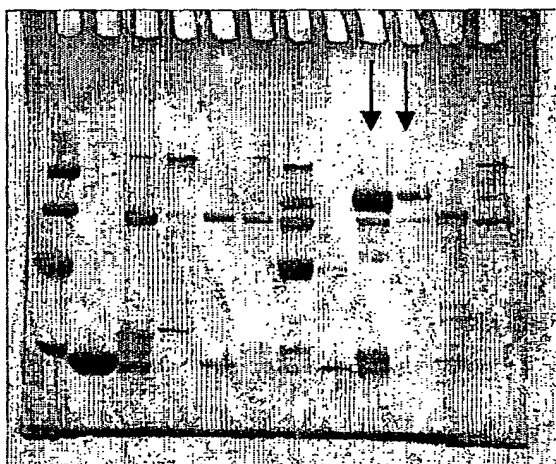
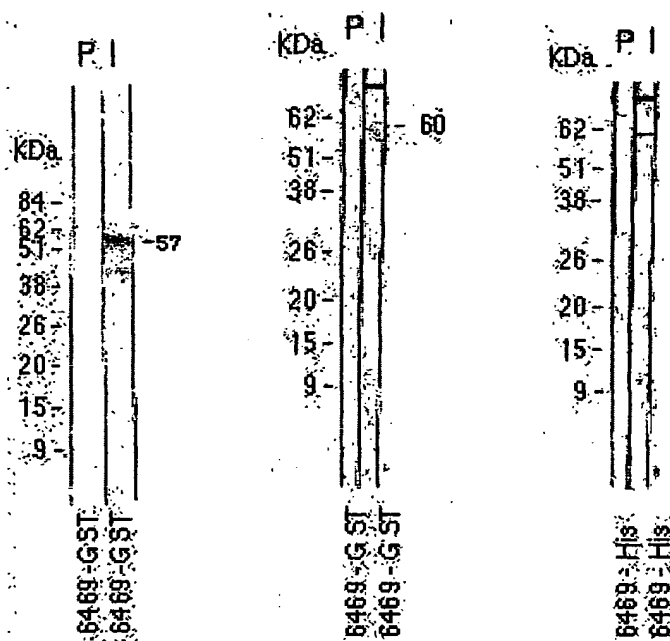
**FIG. 12C**



13/169

**FIGURE 13****Fig. 13A****Fig. 13B**

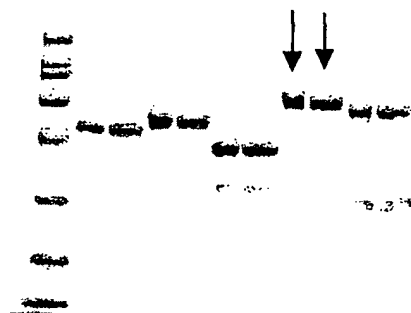
14/169

**FIGURE 14****FIG. 14A****FIG. 14B**

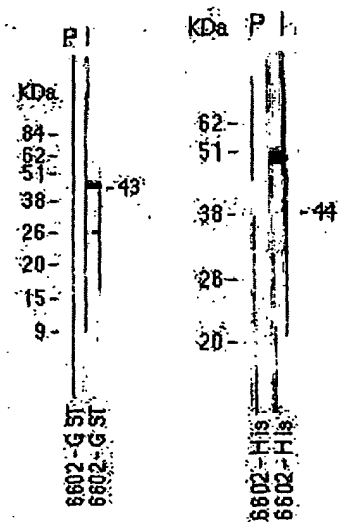
15/169

# **FIGURE 15**

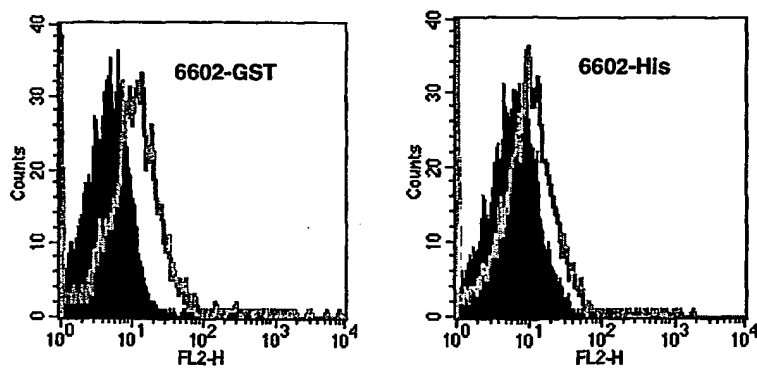
**Fig. 15A**



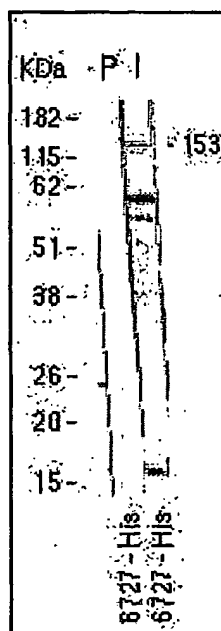
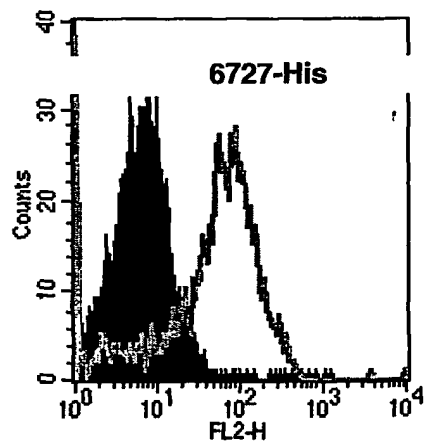
**Fig. 15B**



**Fig. 15C**

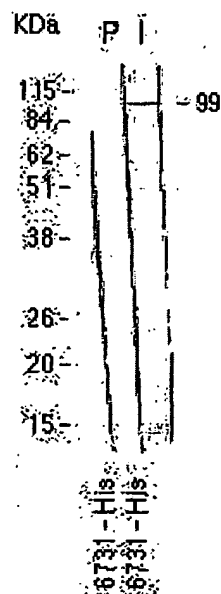
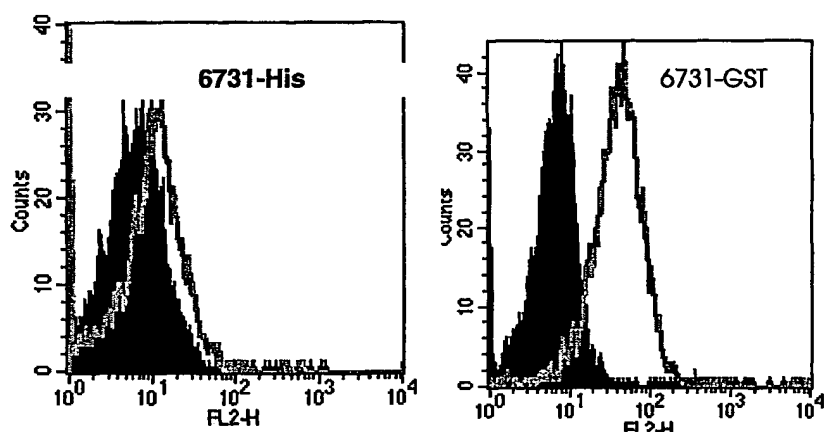


16/169

**FIGURE 16****Fig. 16A****Fig. 16B****Fig. 16C**



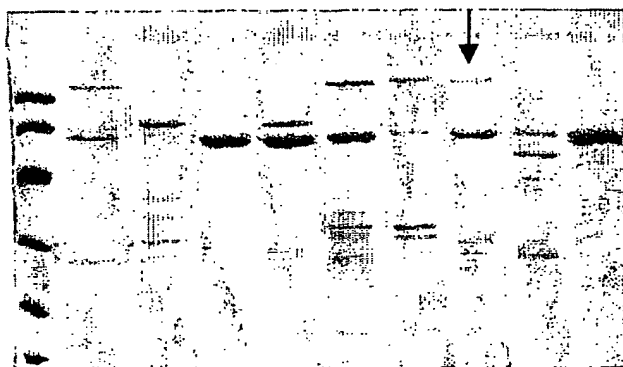
17/169

**FIGURE 17****FIG. 17A****FIG. 17B****FIG. 17C**

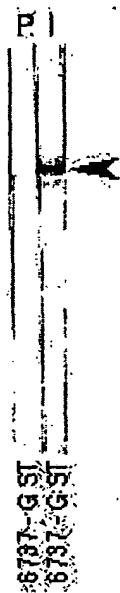
18/169

**FIGURE 18**

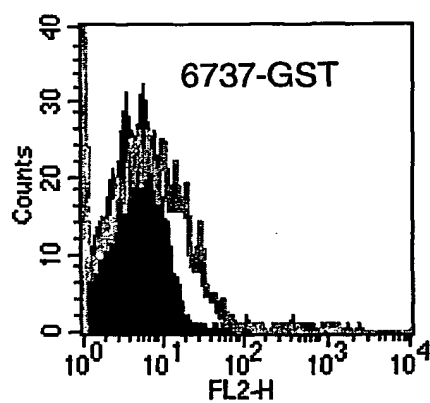
**Fig. 18A**



**Fig. 18B**



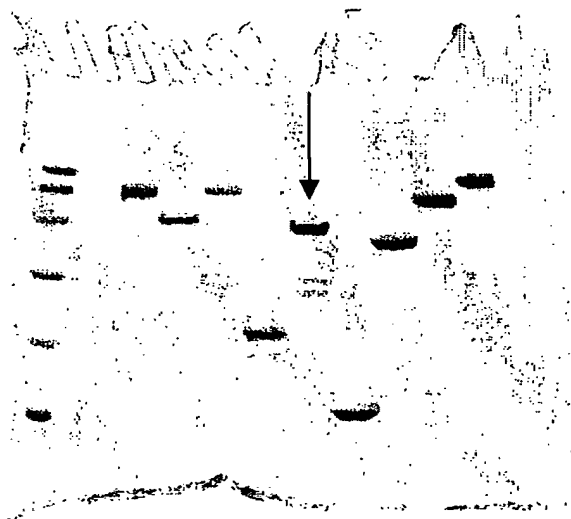
**Fig. 18C**



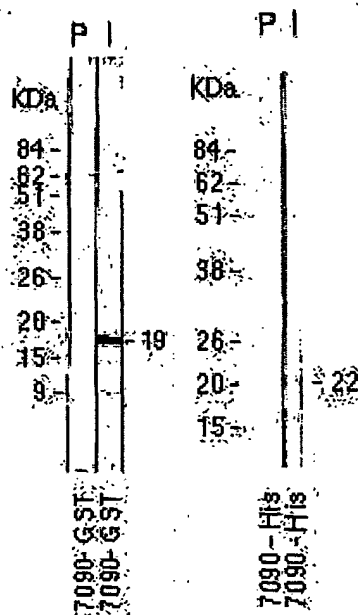
19/169

**FIGURE 19**

**Fig. 19A**



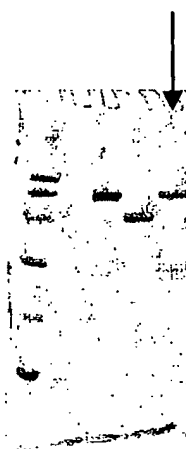
**Fig. 19B**



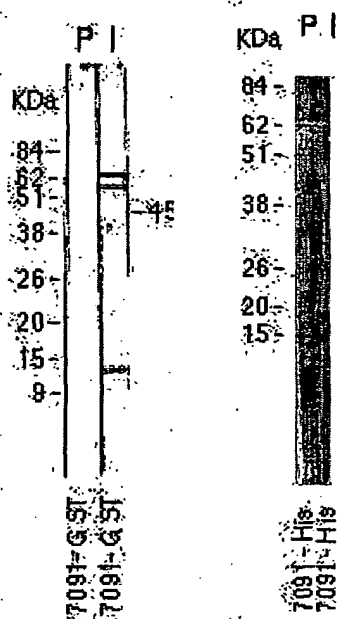
20/169

**FIGURE 20**

**Fig. 20A**



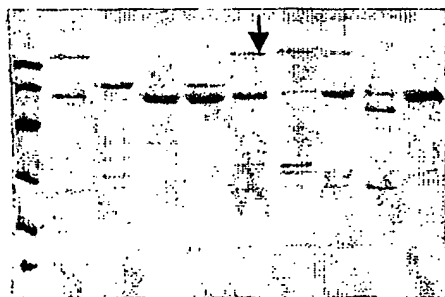
**Fig. 20B**



21/169

**FIGURE 21**

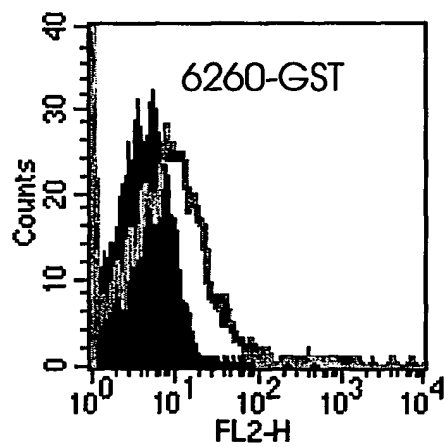
**FIG.  
21A**



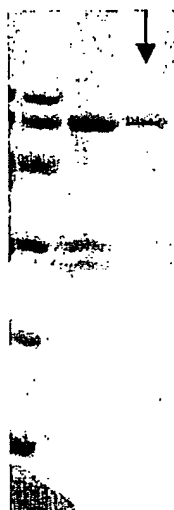
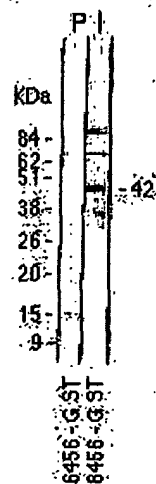
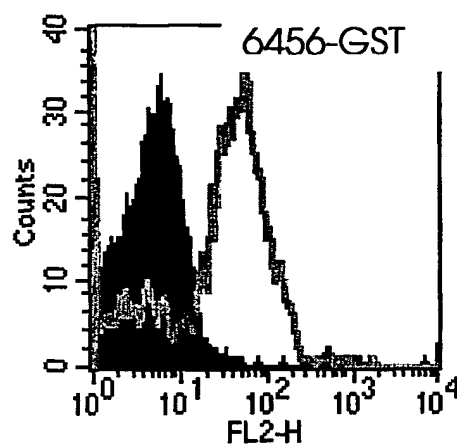
**FIG.  
21B**



**FIG.  
21C**



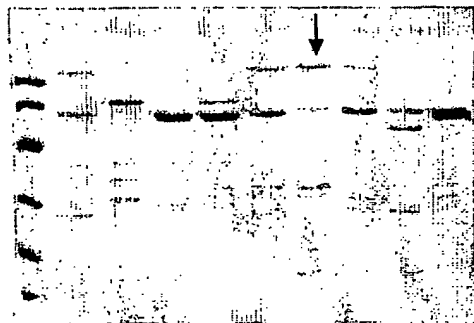
22/169

**FIGURE 22****FIG.  
22A****FIG.  
22B****FIG.  
22C**

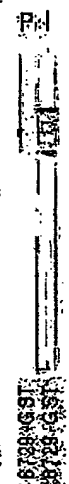
23/169

**FIGURE 23**

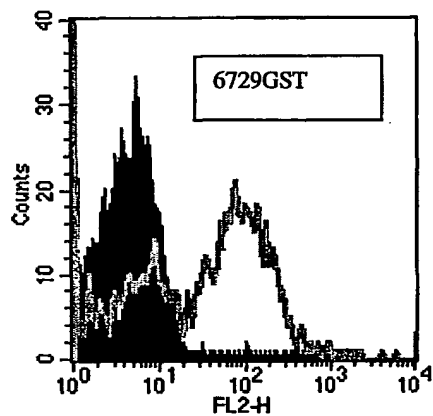
**FIG.  
23A**



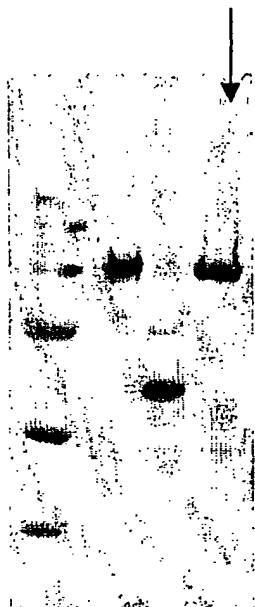
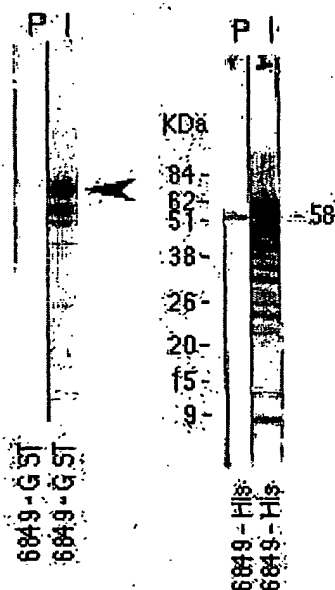
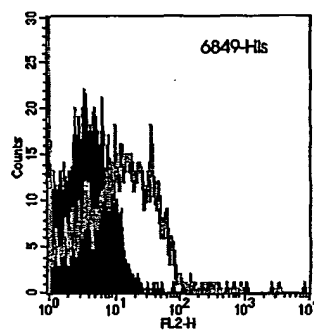
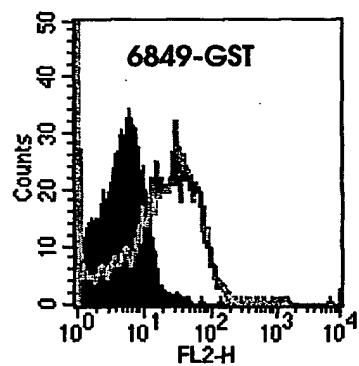
**FIG.  
23B**



**FIG.  
23C**



24/169

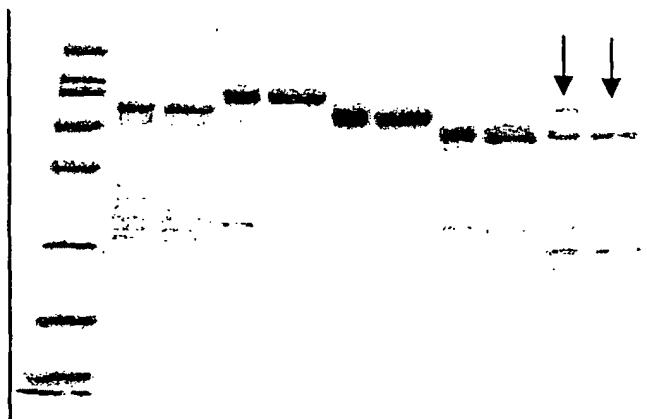
**FIGURE 24****FIG.  
24A****FIG.  
24B****FIG.  
24C**



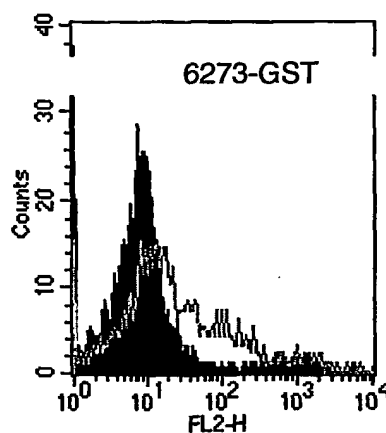
25/169

# **FIGURE 25**

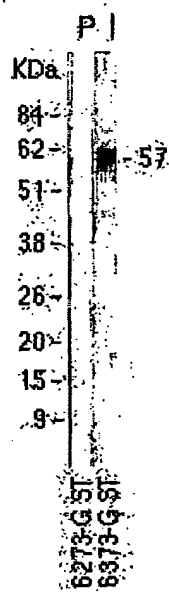
**FIG. 25A**



**Fig. 25C**



**FIG. 25B**



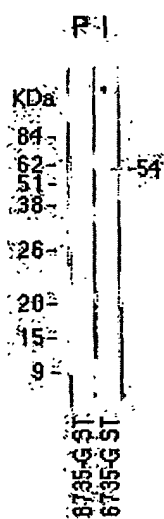
26/169

**FIGURE 26**

**Fig. 26A**

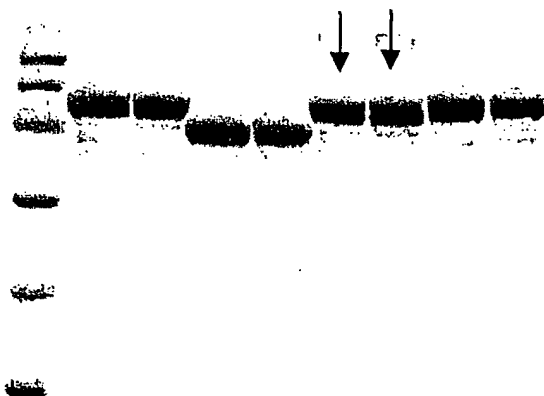


**FIG. 26B**

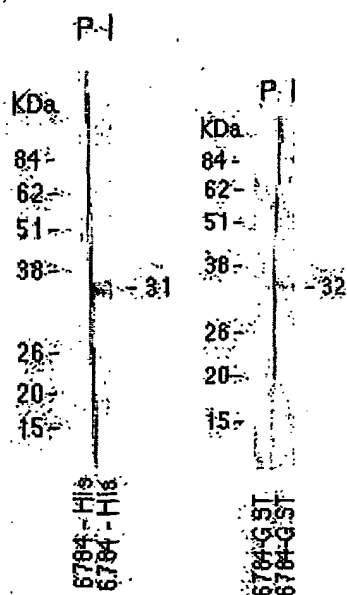


27/169

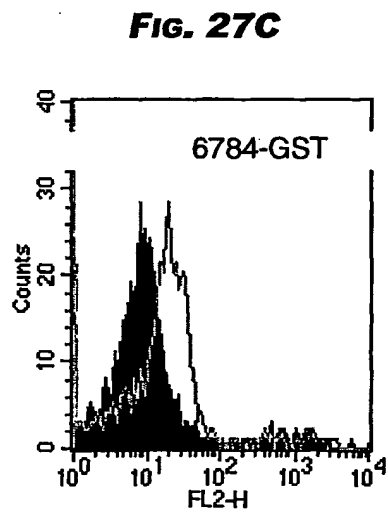
**FIGURE 27**



**Fig. 27A**



**Fig. 27B**

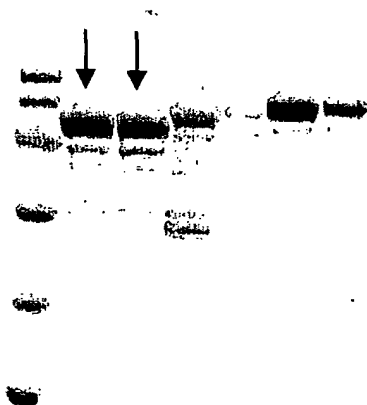


**Fig. 27C**

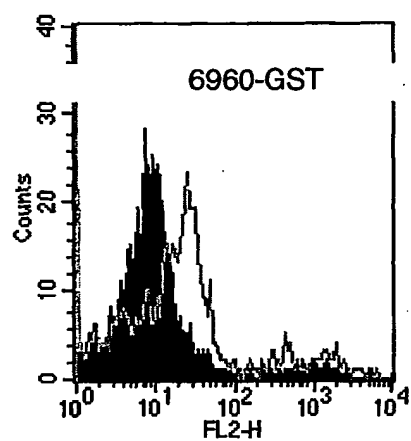
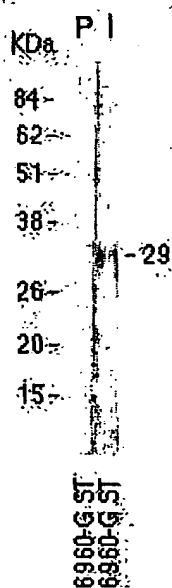
28/169

**FIGURE 28**

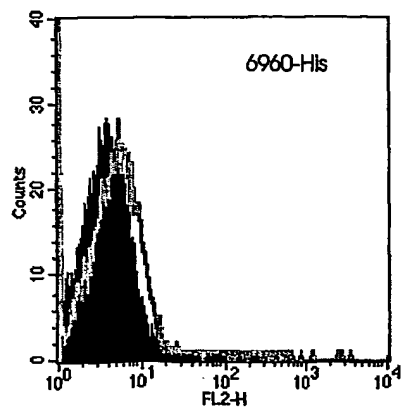
**FIG. 28A**



**FIG. 28B**

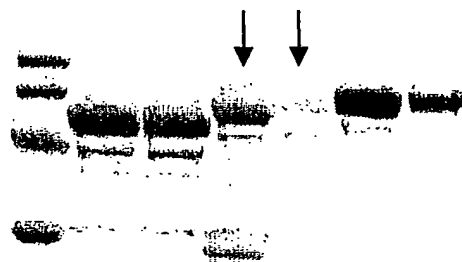


**FIG. 28C**

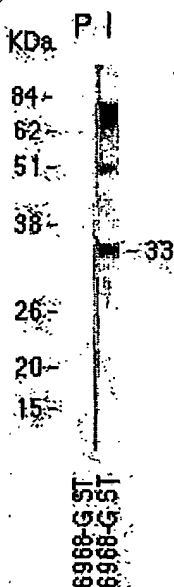


29/169

**FIGURE 29**

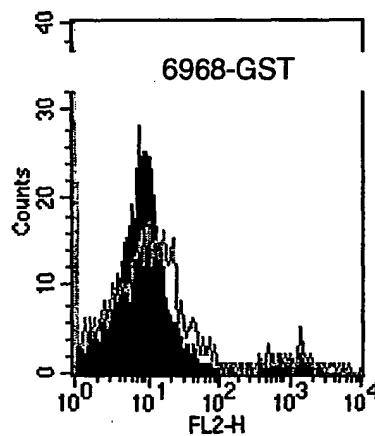


**Fig. 29A**



**Fig. 29B**

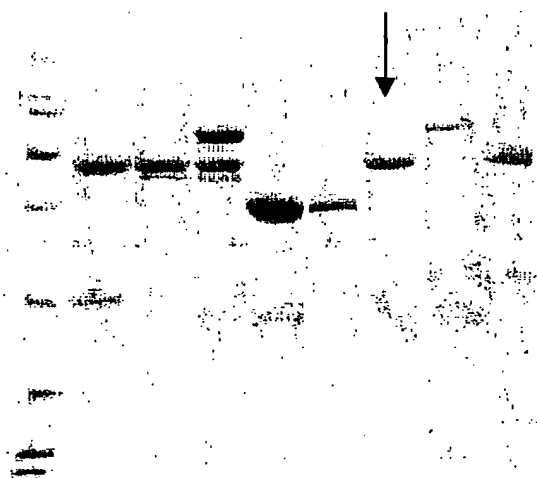
**Fig. 29C**



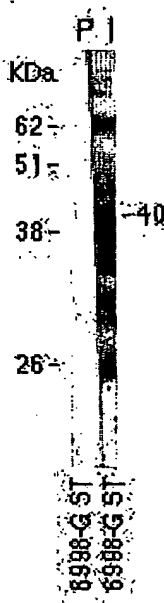
30/169

**FIGURE 30**

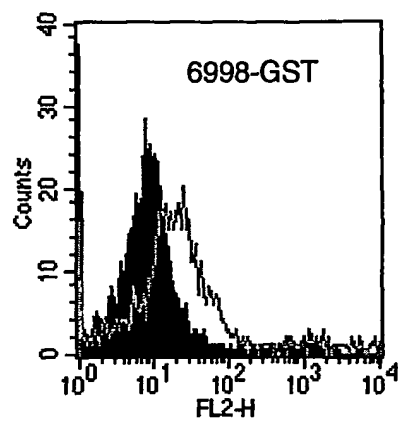
**Fig. 30A**



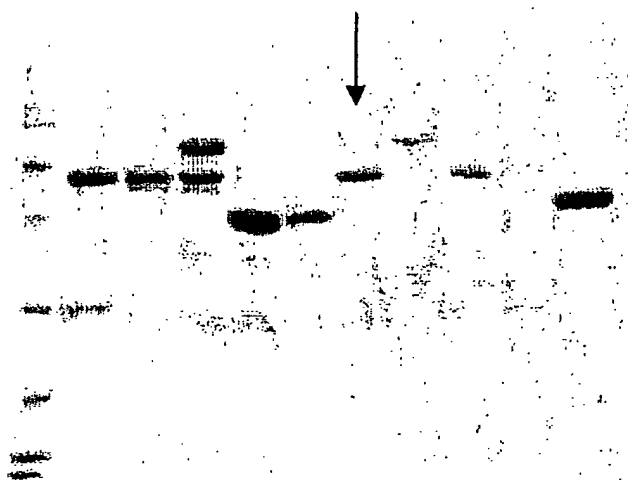
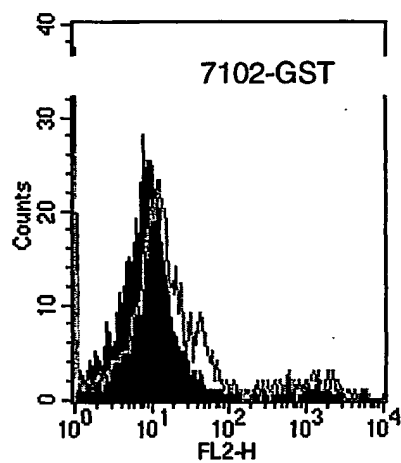
**Fig. 30B**



**Fig. 30C**



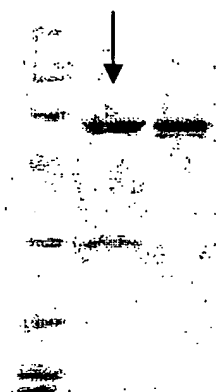
31/169

**FIGURE 31****Fig. 31A****Fig. 31B**

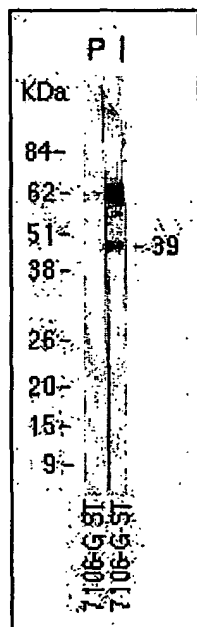
32/169

# **FIGURE 32**

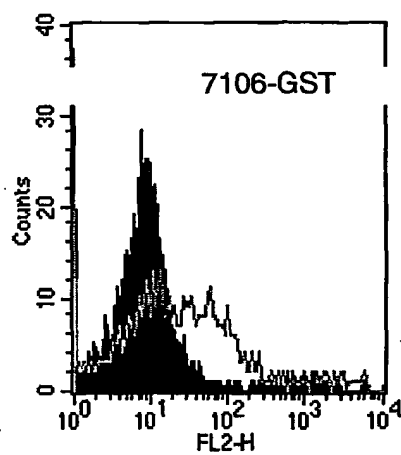
**Fig. 32A**



**Fig. 32B**



**Fig. 32C**

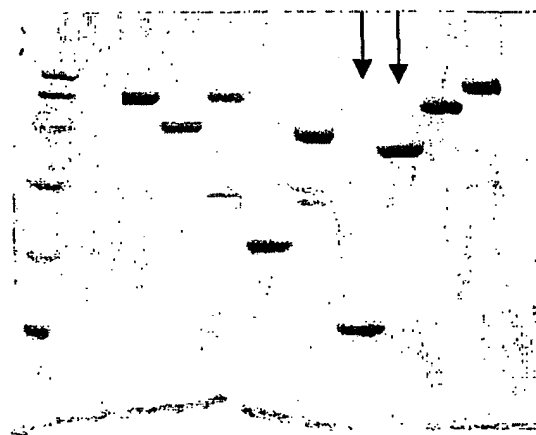




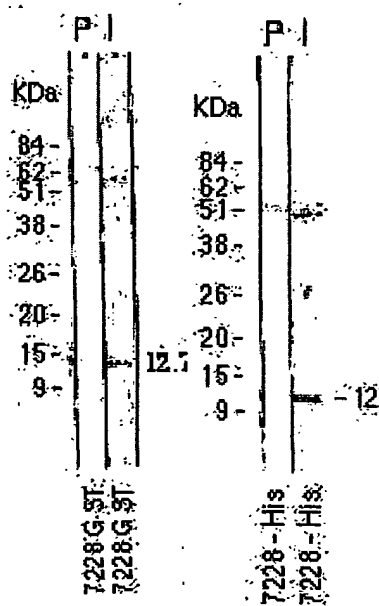
33/169

**FIGURE 33**

**FIG. 33A**



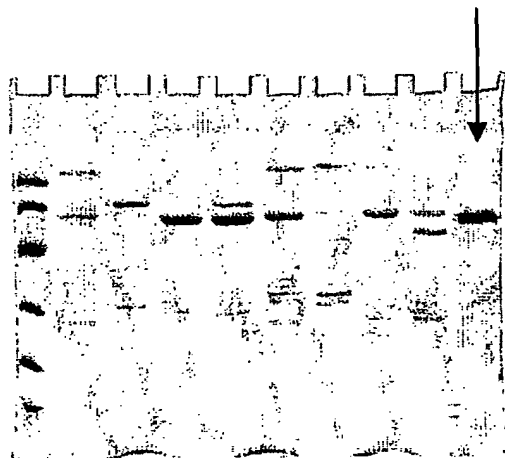
**FIG. 33B**



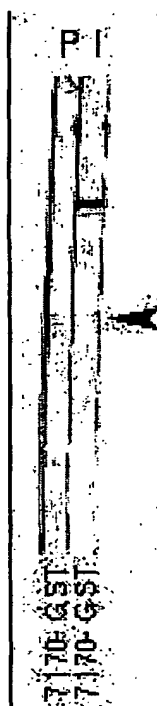
34/169

**FIGURE 34**

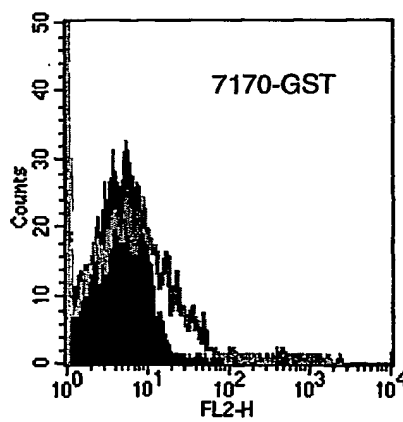
**Fig. 34A**



**Fig. 34B**



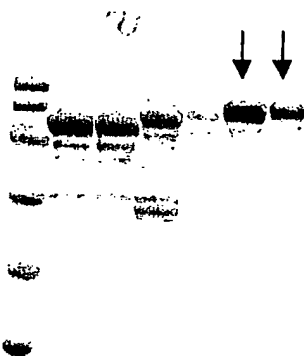
**Fig. 34C**



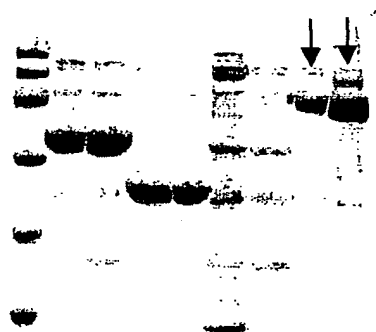
35/169

**FIGURE 35**

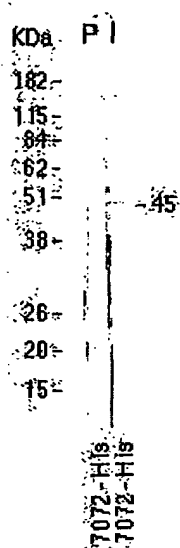
**Fig. 35A**



**Fig. 35B**



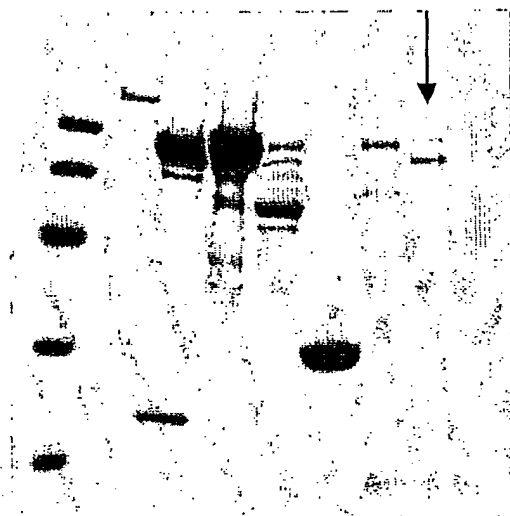
**Fig. 35C**



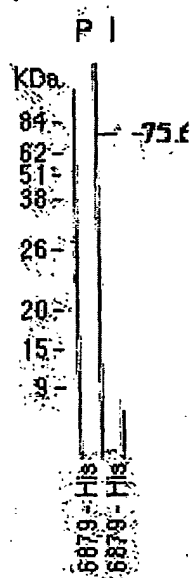
36/169

**FIGURE 36**

**Fig. 36A**



**Fig. 36B**



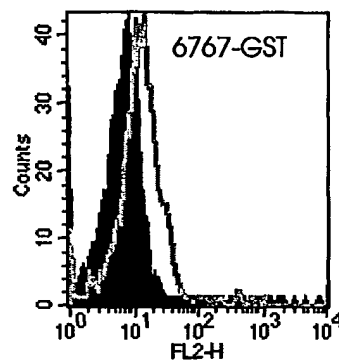
37/169

**FIGURE 37**

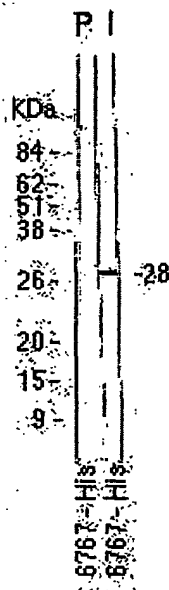
**FIG. 37A**



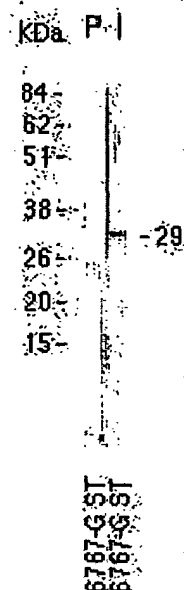
**FIG. 37C**



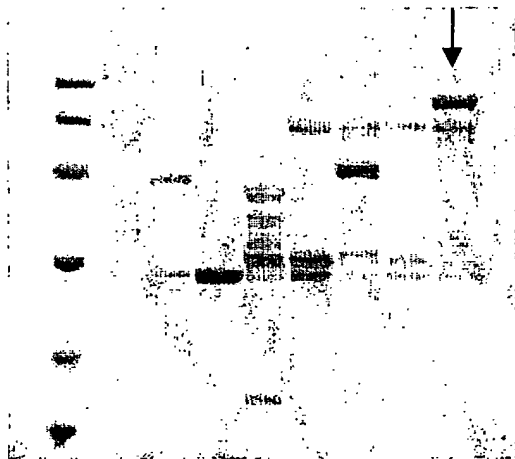
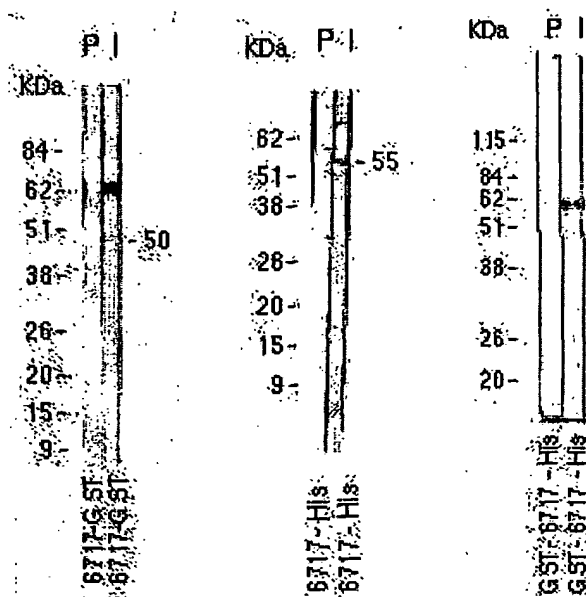
**Fig. 37B**



**Fig. 37D**



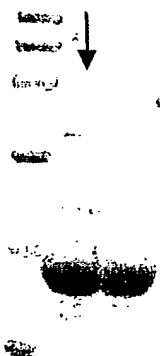
38/169

**FIGURE 38****Fig. 38A****Fig. 38B**

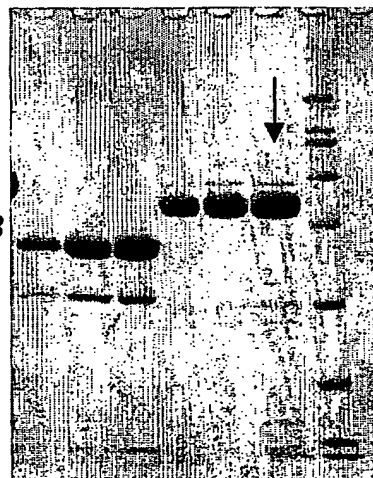
39/169

**FIGURE 39**

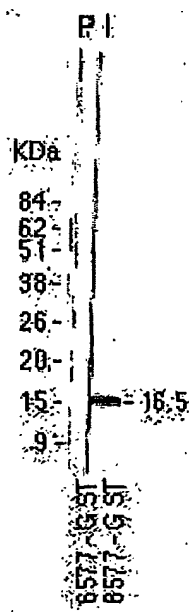
**Fig. 39A**



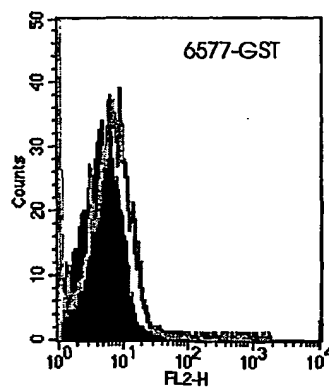
**Fig. 39B**



**Fig. 39C**



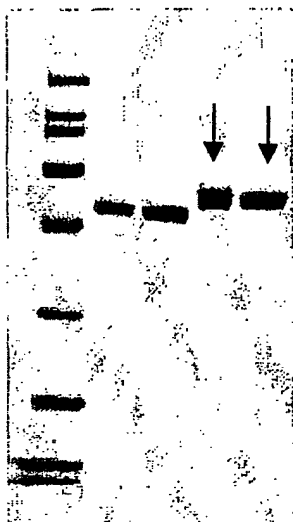
**Fig. 39D**



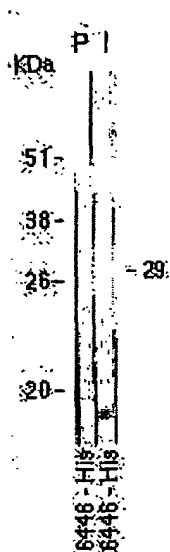
40/169

**FIGURE 40**

**FIG. 40A**



**FIG. 40B**

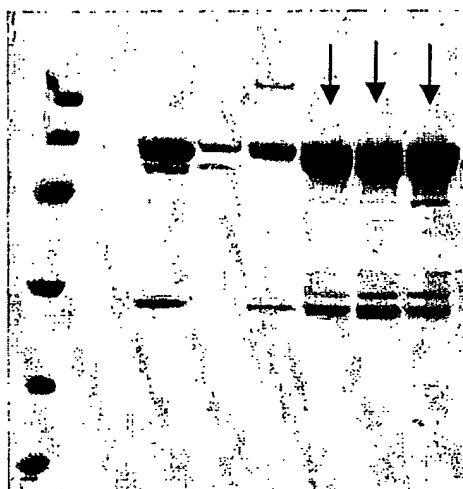




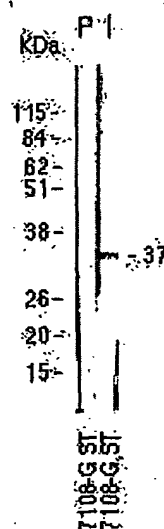
41/169

**FIGURE 41**

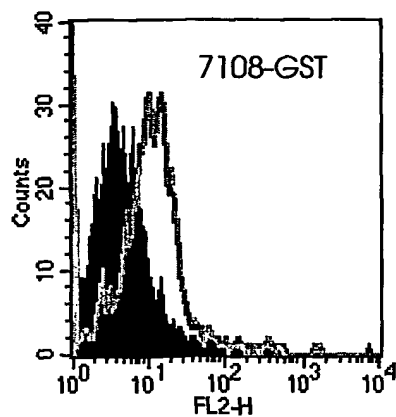
**FIG. 41A**



**FIG. 41B**



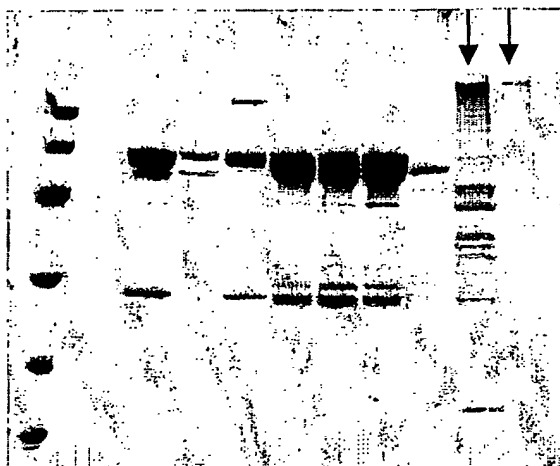
**FIG. 41C**



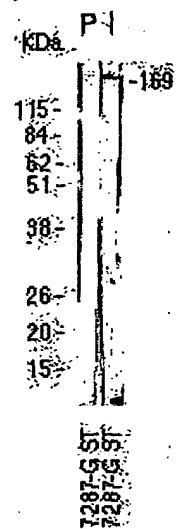
42/169

**FIGURE 42**

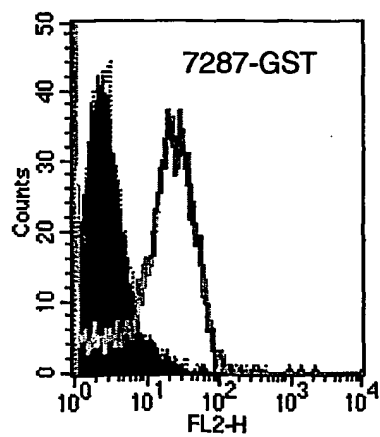
**FIG. 42A**



**FIG. 42B**



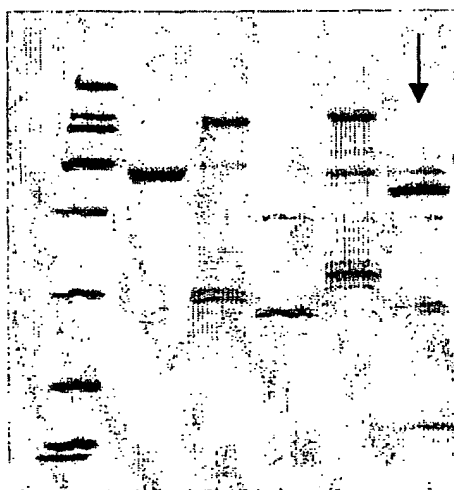
**FIG. 42C**



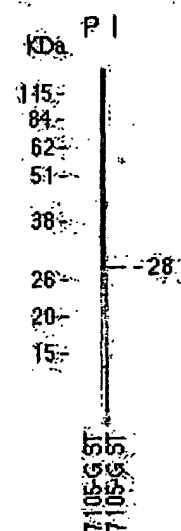
43/169

**FIGURE 43**

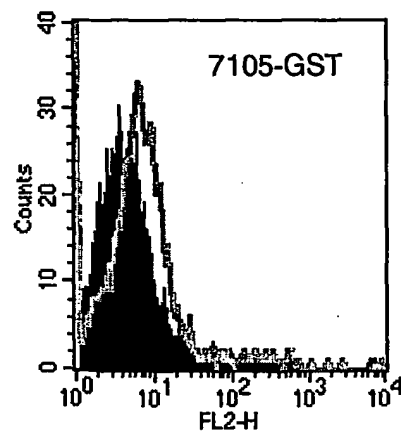
**FIG. 43A**



**FIG. 43B**



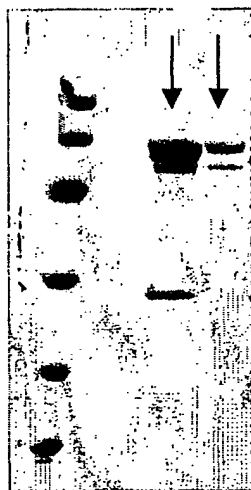
**FIG. 43C**



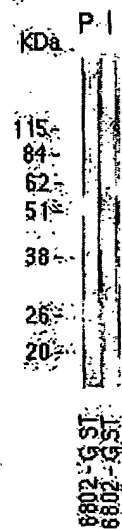
44/169

**FIGURE 44**

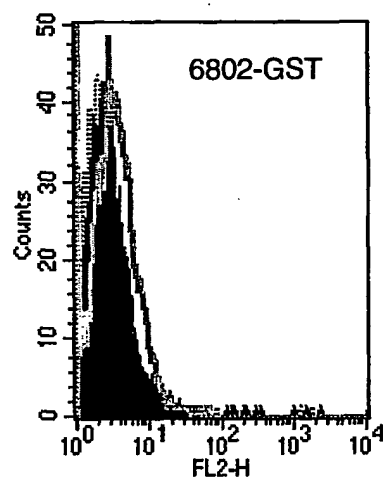
**FIG. 44A**



**FIG. 44B**



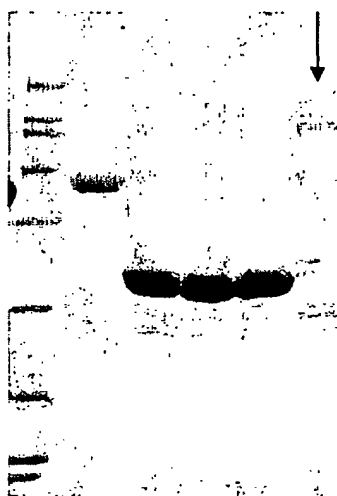
**FIG. 44C**



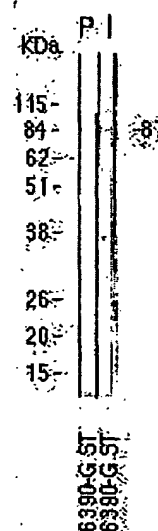
45/169

**FIGURE 45**

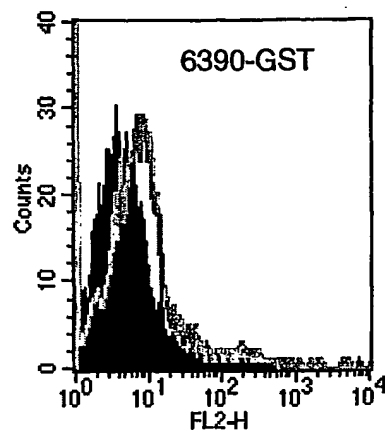
**Fig. 45A**



**Fig. 45B**



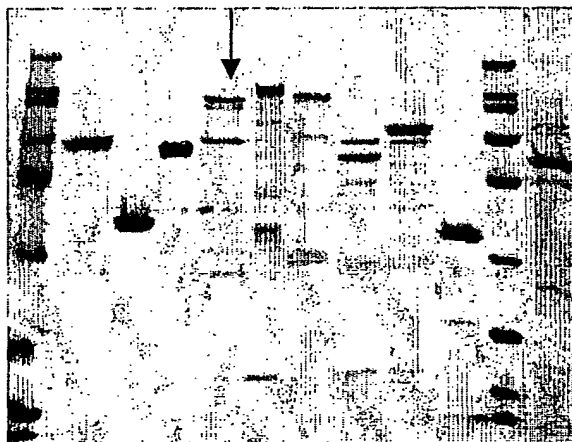
**Fig. 45C**



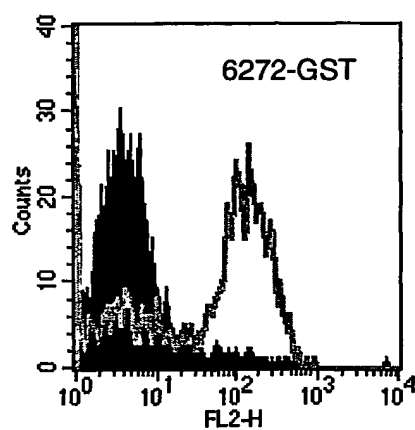
46/169

**FIGURE 46**

**Fig. 46A**



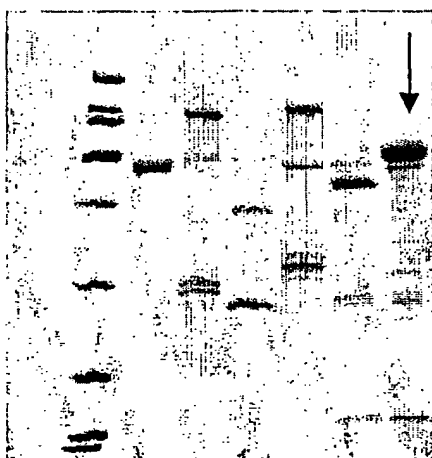
**Fig. 46B**



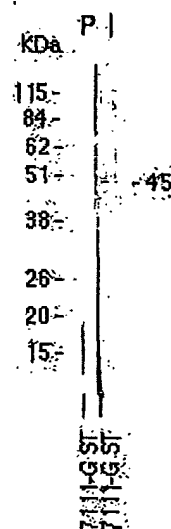
47/169

**FIGURE 47**

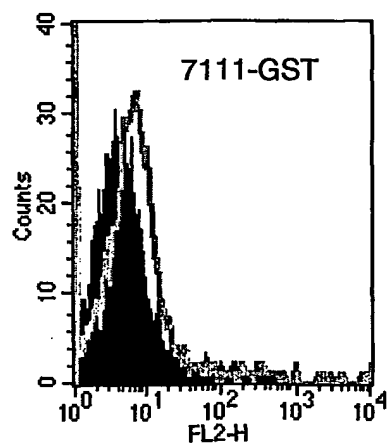
**Fig. 47A**



**FIG. 47B**



**Fig. 47C**



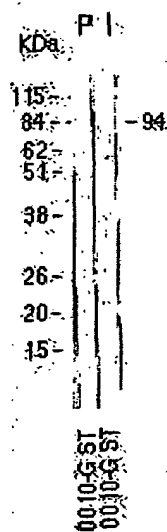
48/169

**FIGURE 48**

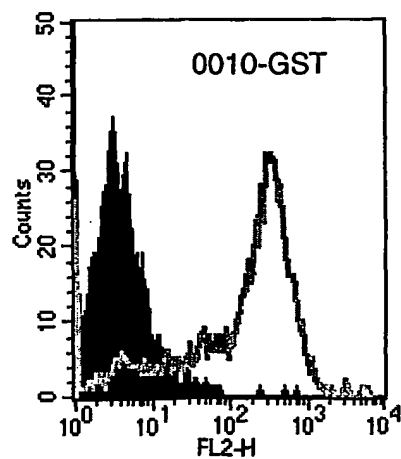
**Fig. 48A**



**Fig. 48B**



**Fig. 48C**

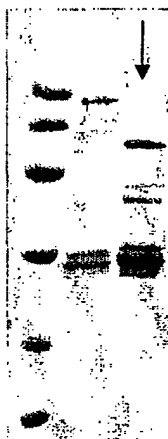




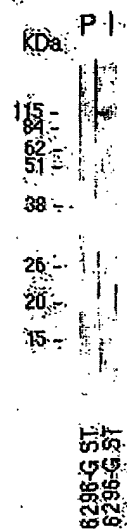
49/169

**FIGURE 49**

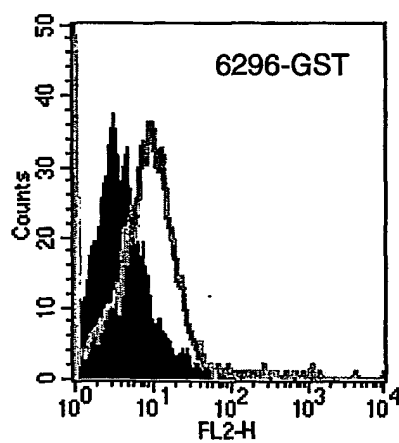
**Fig. 49A**



**Fig. 49B**



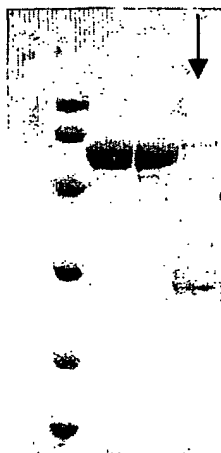
**Fig. 49C**



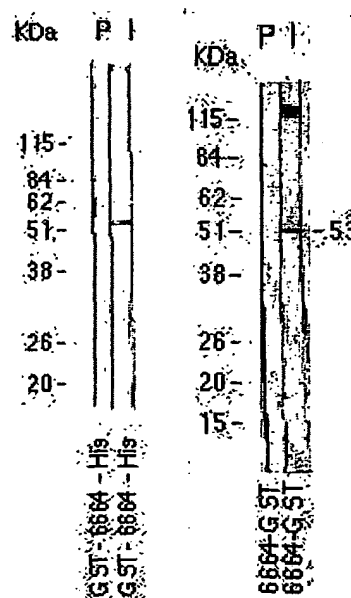
50/169

**FIGURE 50**

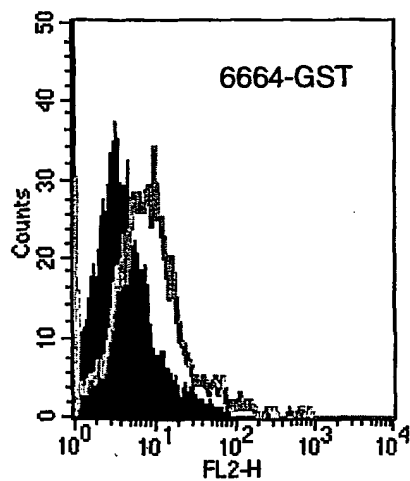
**Fig. 50A**



**Fig. 50B**



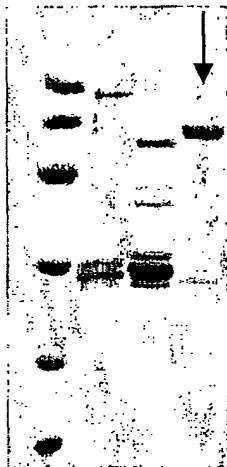
**Fig. 50C**



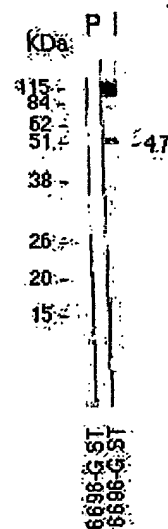
51/169

**FIGURE 51**

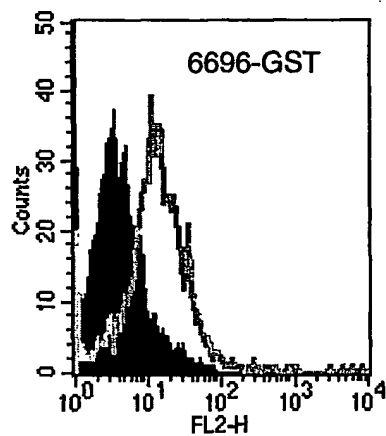
**Fig. 51A**



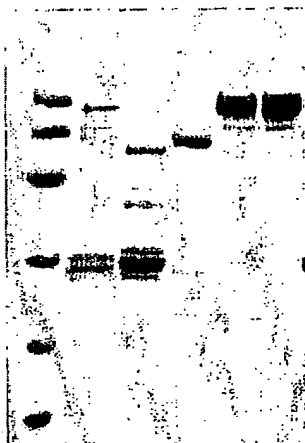
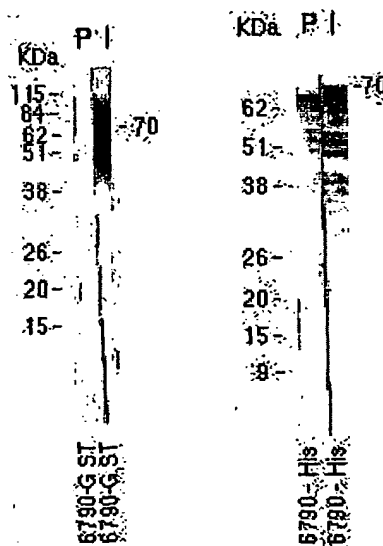
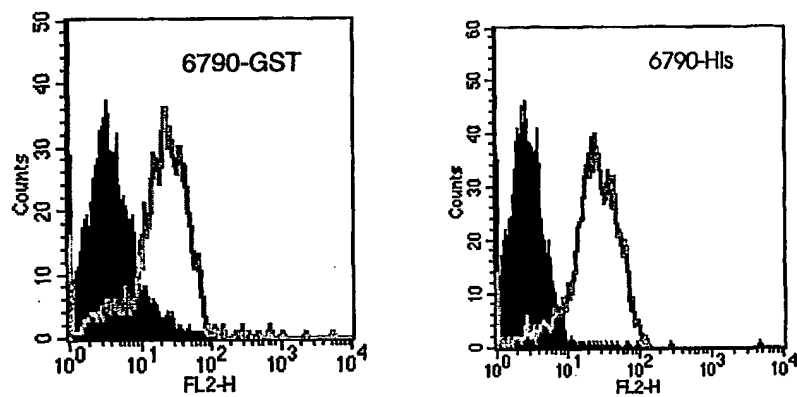
**Fig. 51B**



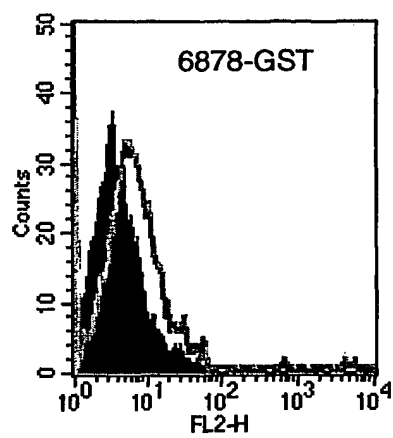
**Fig. 51C**



52/169

**FIGURE 52****Fig. 52A****Fig. 52B****Fig. 52C**

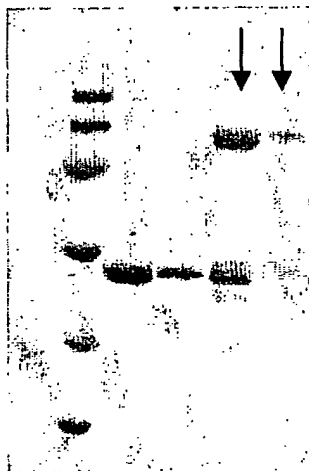
53/169

**FIGURE 53****Fig. 53A****Fig. 53B**

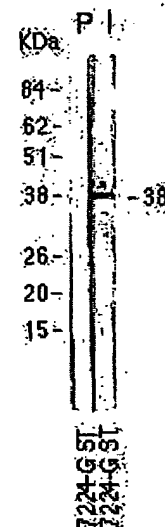
54/169

**FIGURE 54**

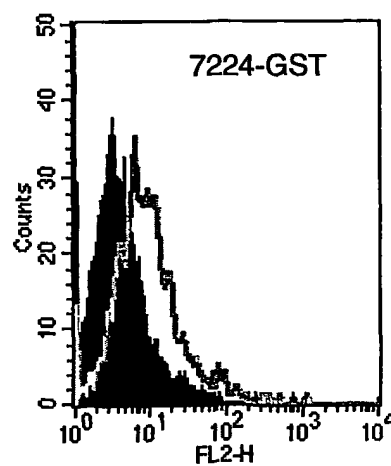
**Fig. 54A**



**Fig. 54B**



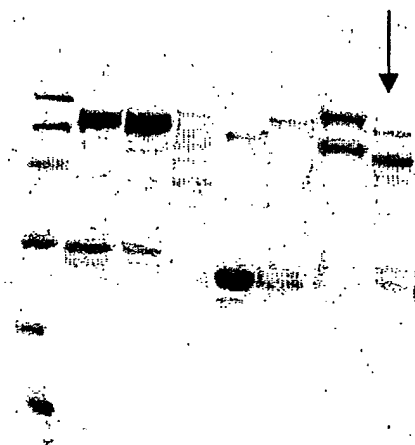
**Fig. 54C**



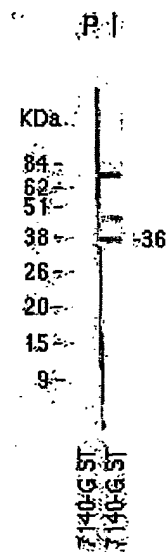
55/169

**FIGURE 55**

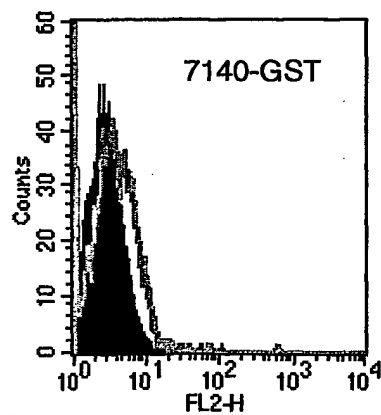
**Fig. 55A**



**Fig. 55B**



**Fig. 55C**



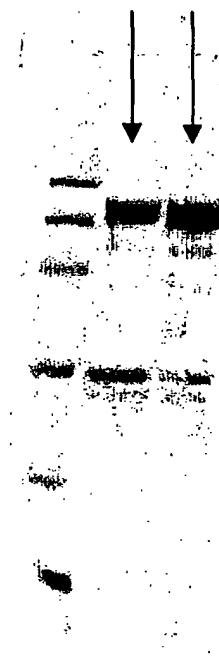
56/169

**FIGURE 56**

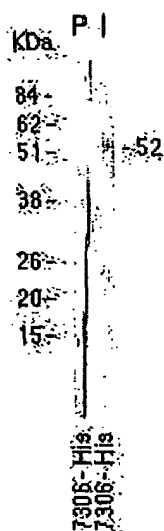
**Fig. 56A**



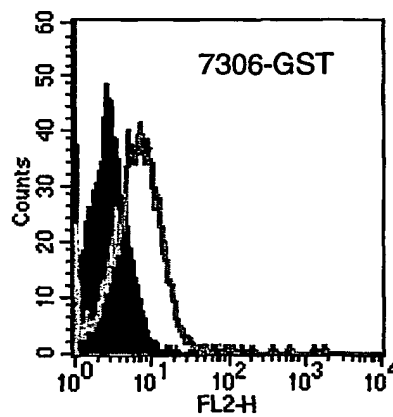
**Fig. 56B**



**FIG. 56C**



**FIG. 56D**

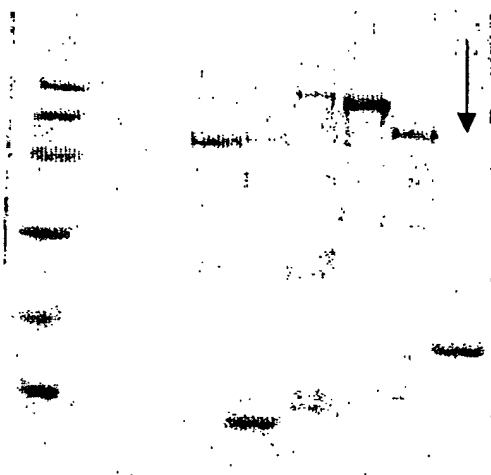




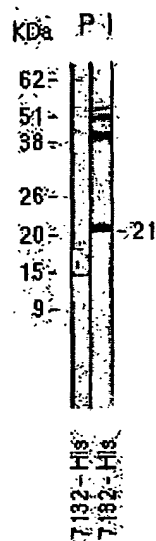
57/169

**FIGURE 57**

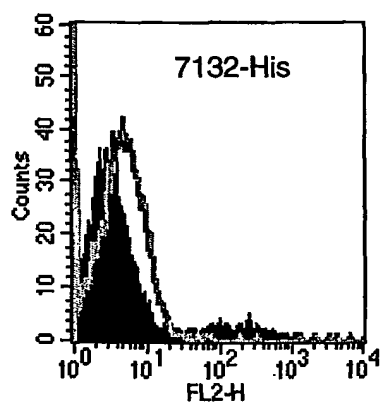
**Fig. 57A**



**Fig. 57B**



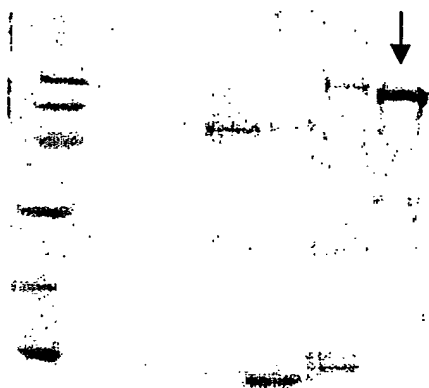
**Fig. 57C**



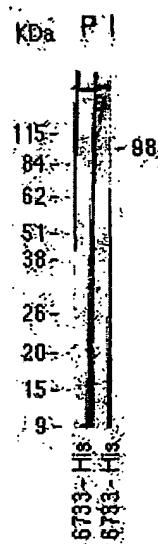
58/169

**FIGURE 58**

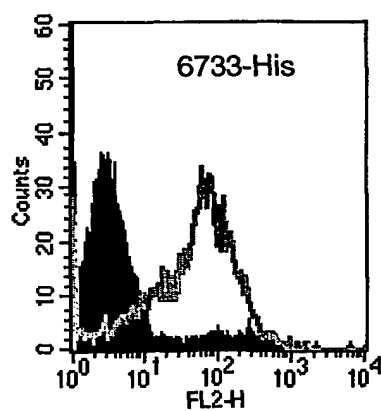
**FIG. 58A**



**FIG. 58B**



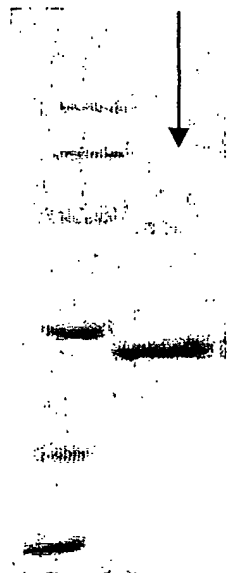
**FIG. 58C**



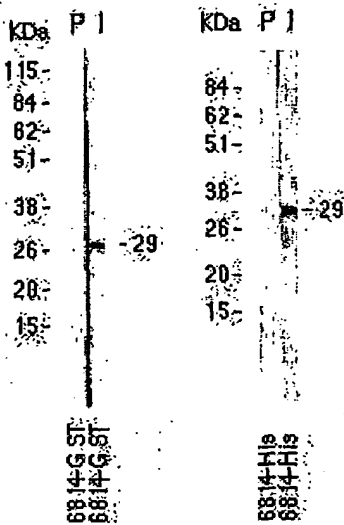
59/169

**FIGURE 59**

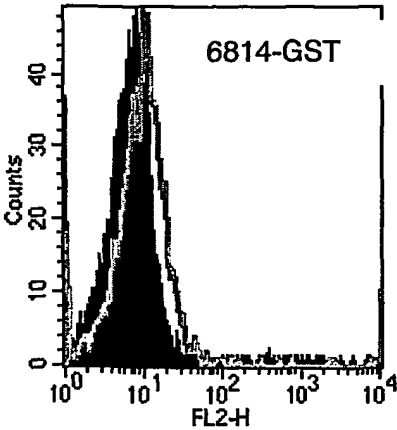
**Fig. 59A**



**Fig. 59B**



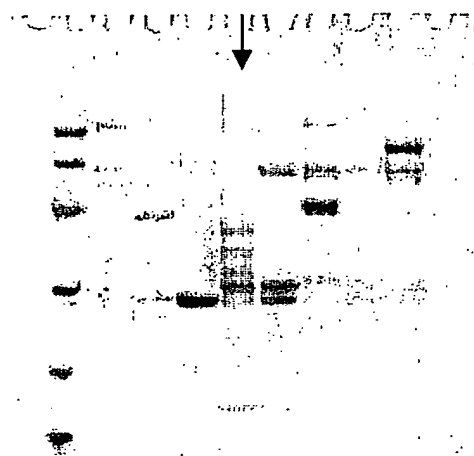
**Fig. 59C**



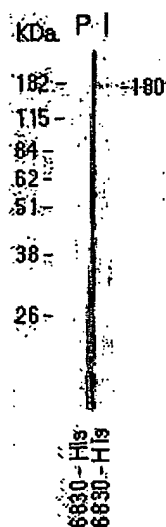
60/169

**FIGURE 60**

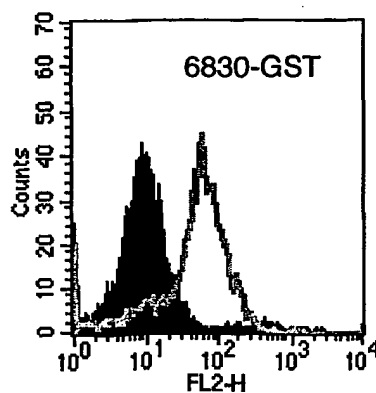
**FIG. 60A**



**FIG. 60B**



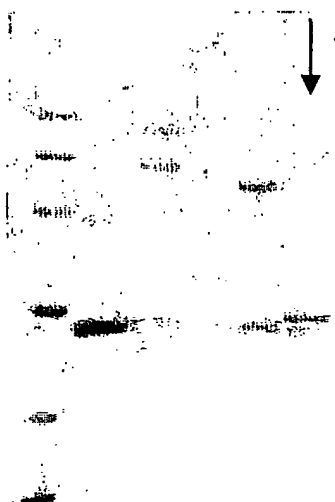
**FIG. 60C**



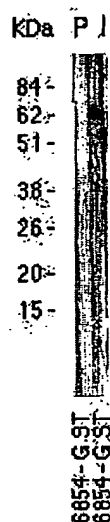
61/169

**FIGURE 61**

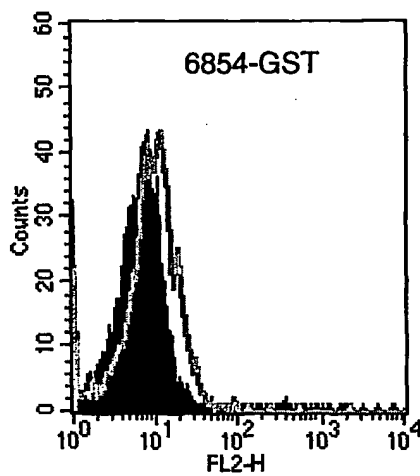
**FIG. 61A**



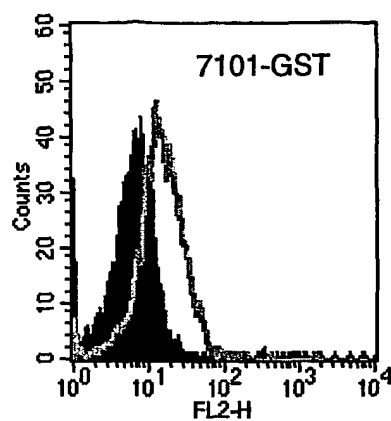
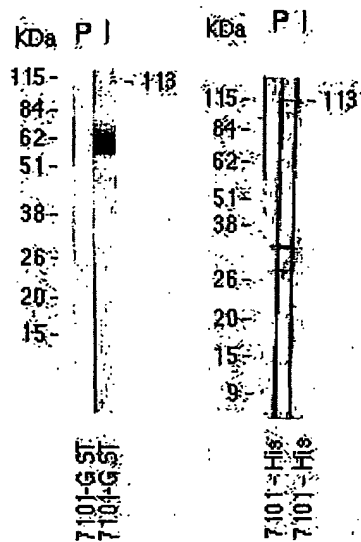
**FIG. 61B**



**FIG. 61C**



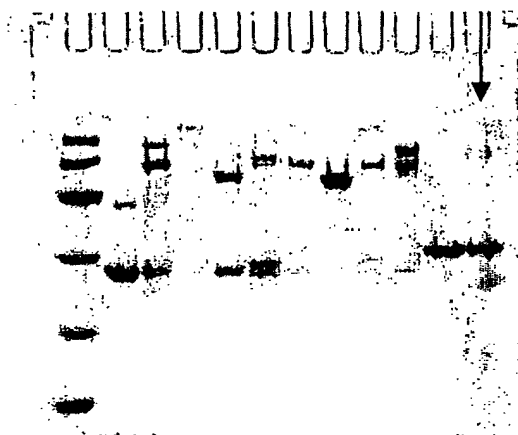
62/169

**FIGURE 62****Fig. 62A****Fig. 62C****Fig. 62B**

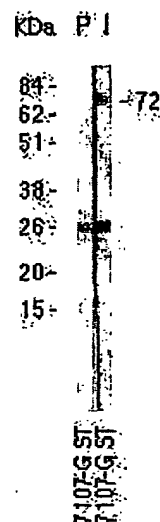
63/169

**FIGURE 63**

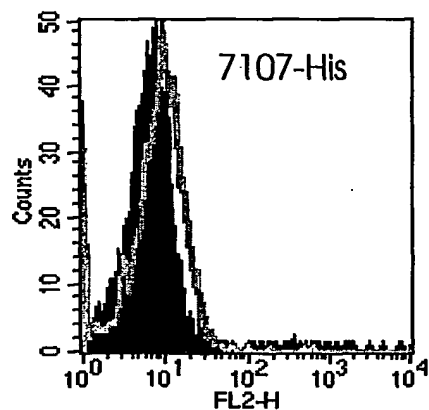
**FIG. 63A**



**FIG. 63B**



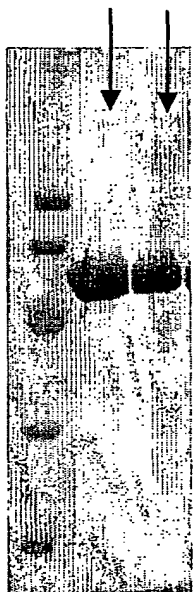
**FIG. 63C**



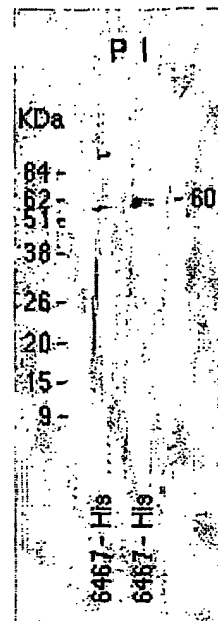
64/169

**FIGURE 64**

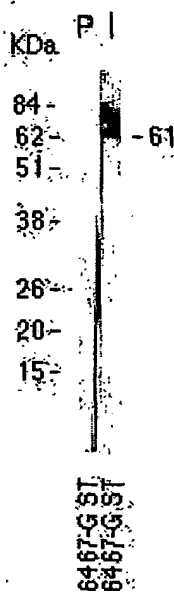
**FIG. 64A**



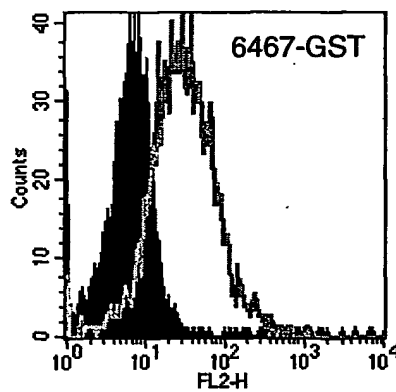
**FIG. 64B**



**FIG. 64C**



**FIG. 64D**

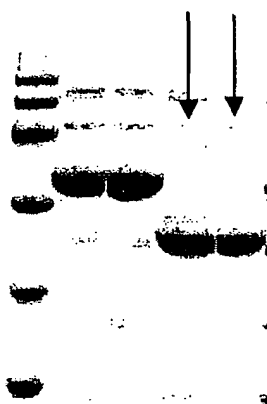




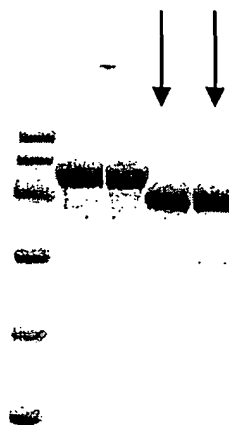
65/169

**FIGURE 65**

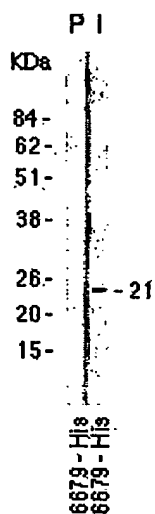
**Fig. 65A**



**Fig. 65B**



**Fig. 65C**

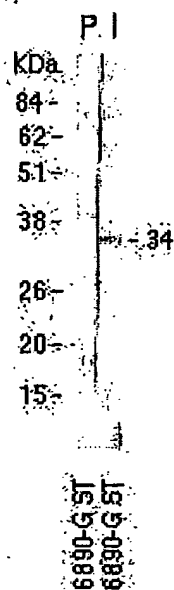


66/169

**FIGURE 66**



**Fig. 66A**



**Fig. 66B**

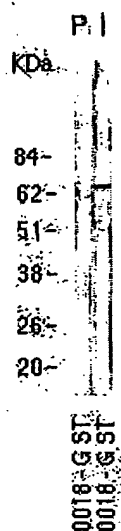
67/169

**FIGURE 67**

**Fig. 67A**



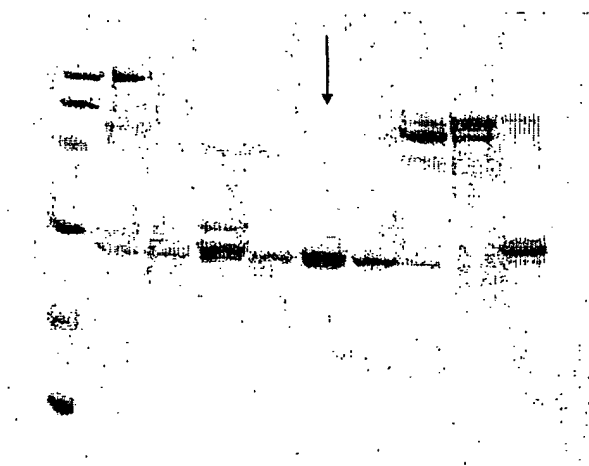
**Fig. 67B**



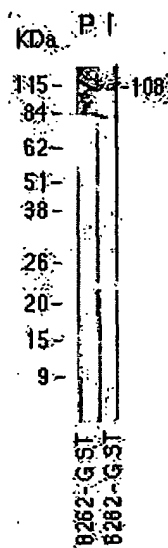
68/169

**FIGURE 68**

**Fig. 68A**



**Fig. 68B**



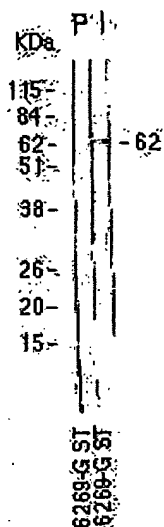
69/169

**FIGURE 69**

**Fig. 69A**



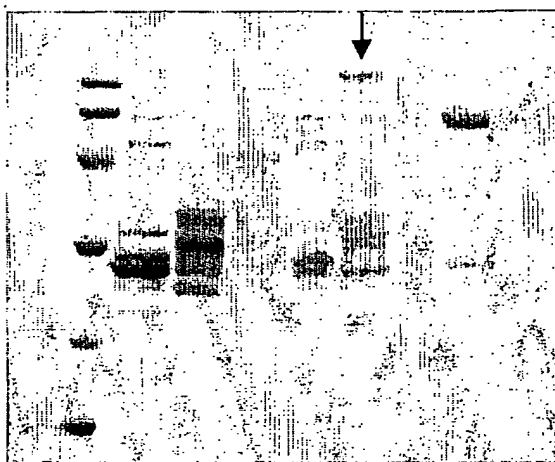
**Fig. 69B**



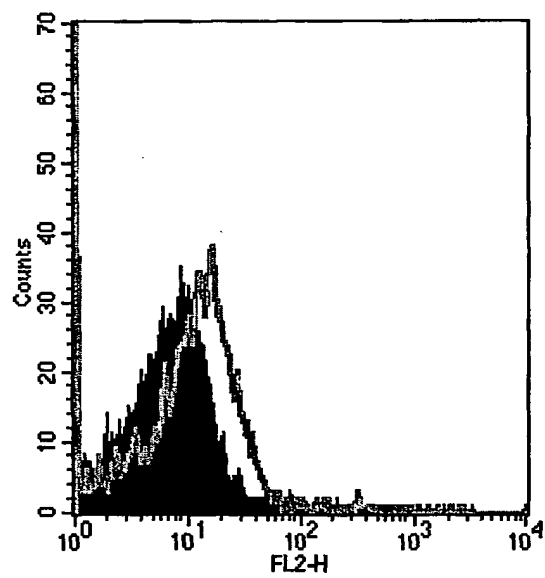
70/169

**FIGURE 70**

**Fig. 70A**



**Fig. 70B**



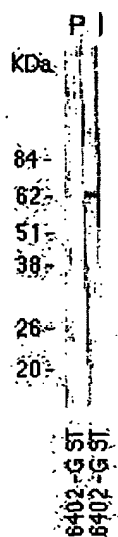
71/169

**FIGURE 71**

**FIG. 71A**



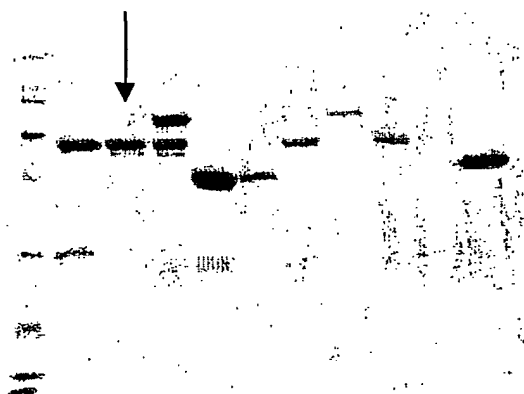
**FIG. 71B**



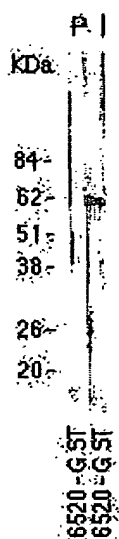
72/169

**FIGURE 72**

**FIG. 72A**



**FIG. 72B**

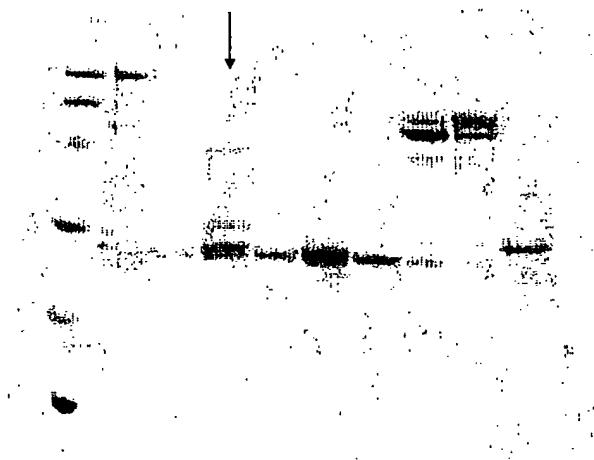




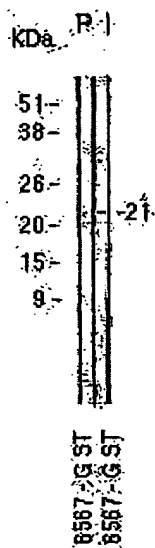
73/169

**FIGURE 73**

**Fig. 73A**



**Fig. 73B**



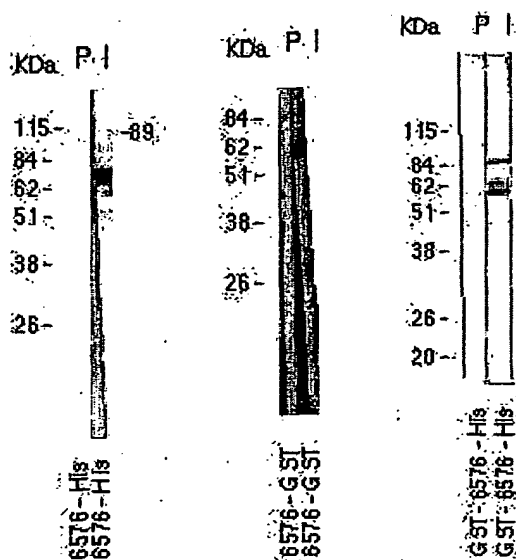
74/169

**FIGURE 74**

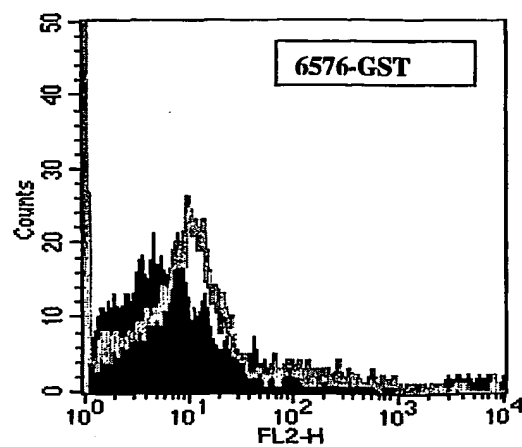
**Fig. 74A**



**Fig. 74B**



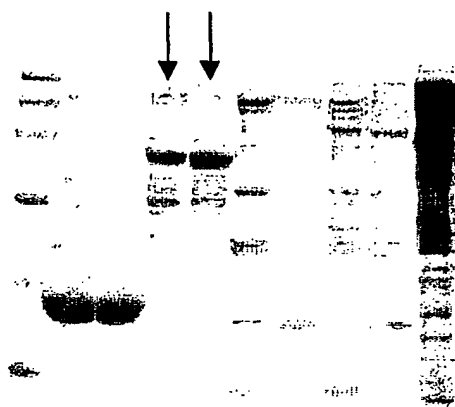
**Fig. 74C**



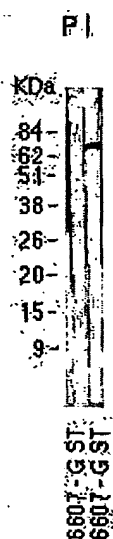
75/169

**FIGURE 75**

**Fig. 75A**



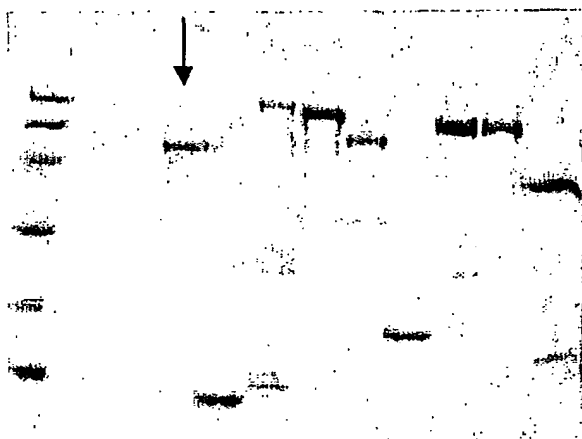
**Fig. 75B**



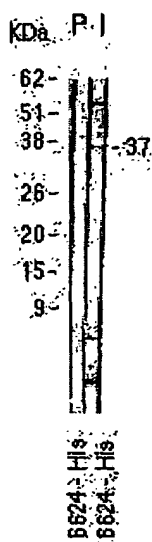
76/169

**FIGURE 76**

**Fig. 76A**



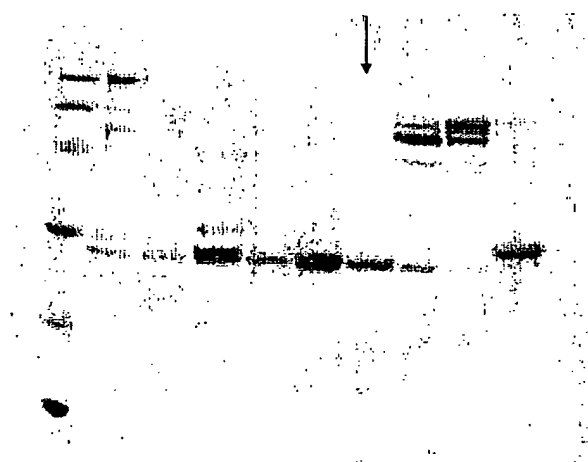
**Fig. 76B**



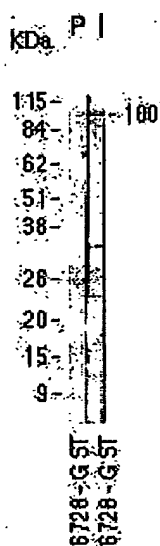
77/169

**FIGURE 77**

**Fig. 77A**



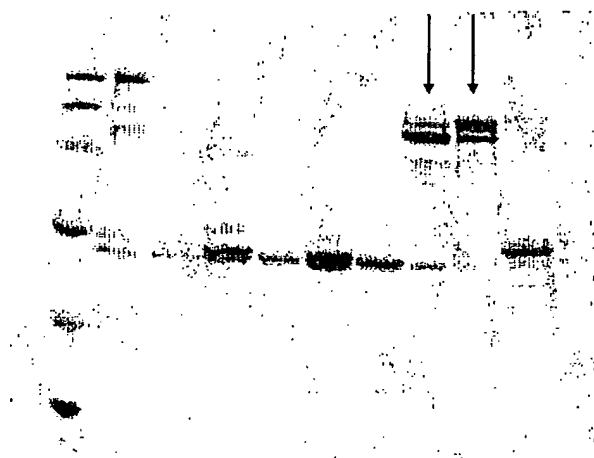
**Fig. 77B**



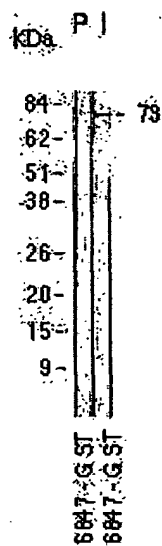
78/169

**FIGURE 78**

**Fig. 78A**



**Fig. 78B**



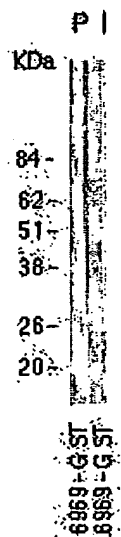
79/169

**FIGURE 79**

**Fig. 79A**



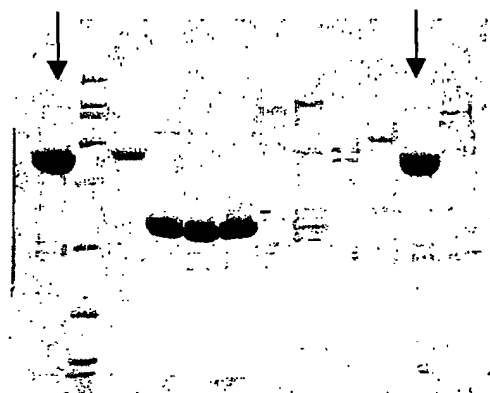
**Fig. 79B**



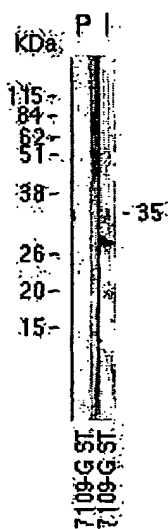
80/169

**FIGURE 80**

**Fig. 80A**



**Fig. 80B**

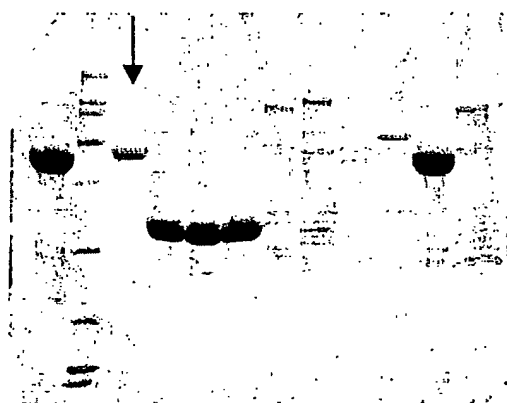




81/169

**FIGURE 81**

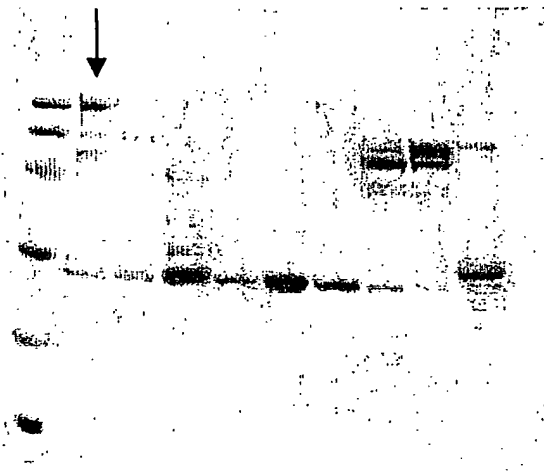
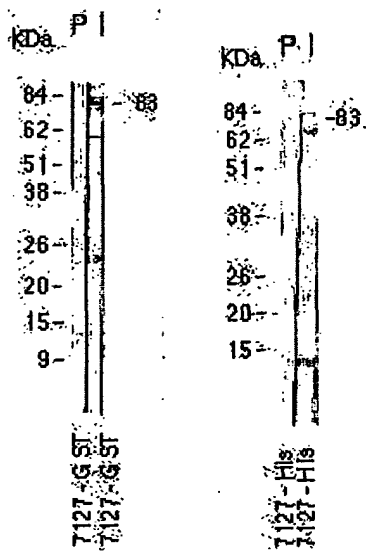
**Fig. 81A**



**Fig. 81B**



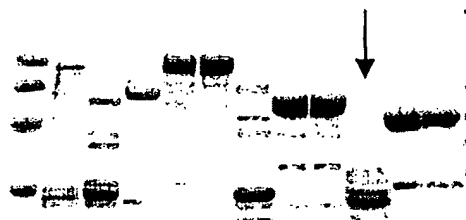
82/169

**FIGURE 82****Fig. 82A****Fig. 82B**

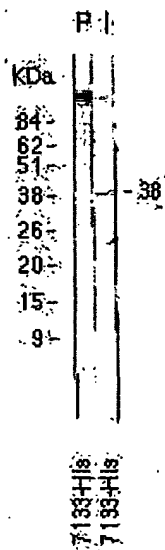
83/169

**FIGURE 83**

**Fig. 83A**



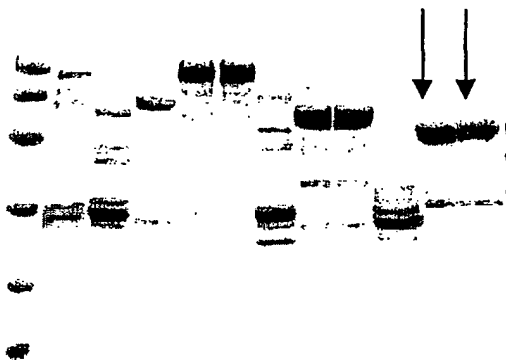
**Fig. 83B**



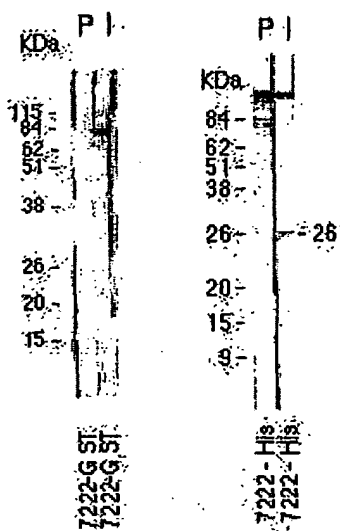
84/169

**FIGURE 84**

**Fig. 84A**



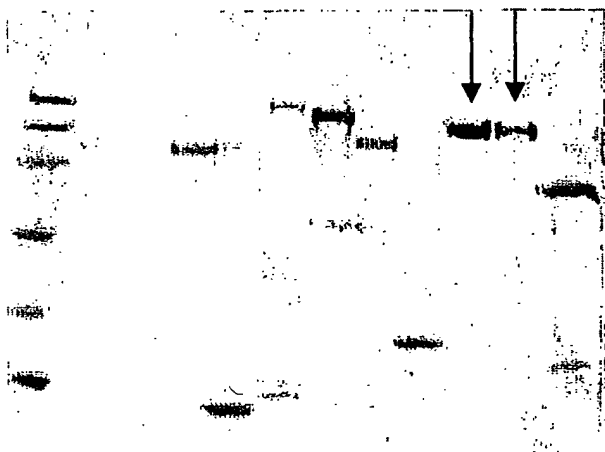
**Fig. 84B**



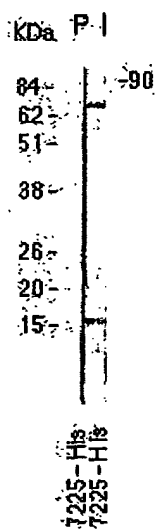
85/169

**FIGURE 85**

**Fig. 85A**



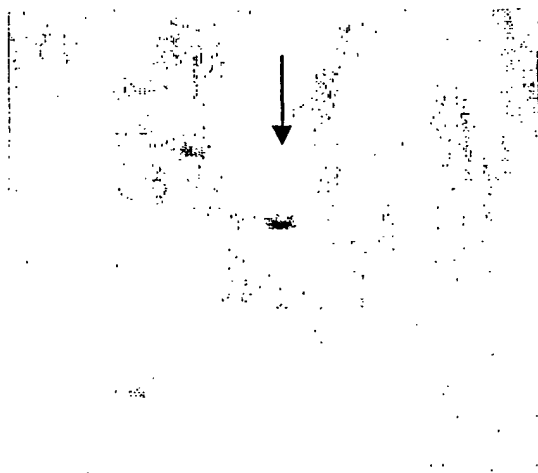
**Fig. 85B**



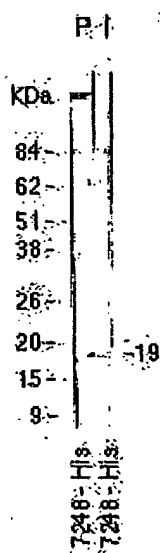
86/169

**FIGURE 86**

**Fig. 86A**



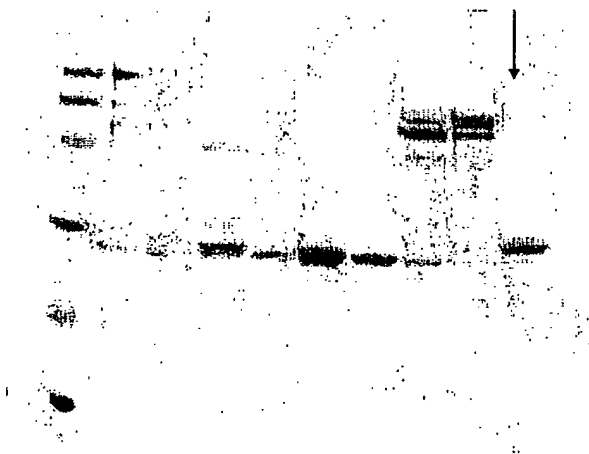
**Fig. 86B**



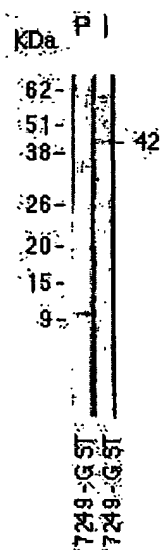
87/169

**FIGURE 87**

**Fig. 87A**



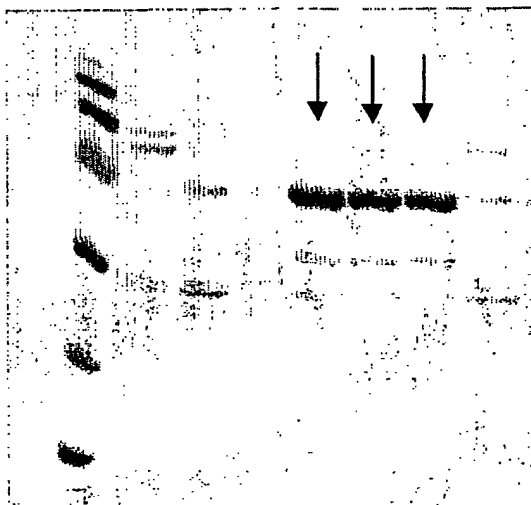
**Fig. 87B**



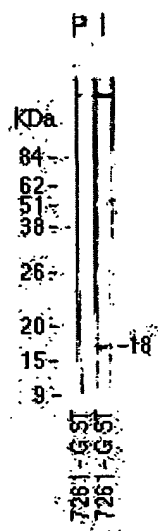
88/169

**FIGURE 88**

**Fig. 88A**



**Fig. 88B**

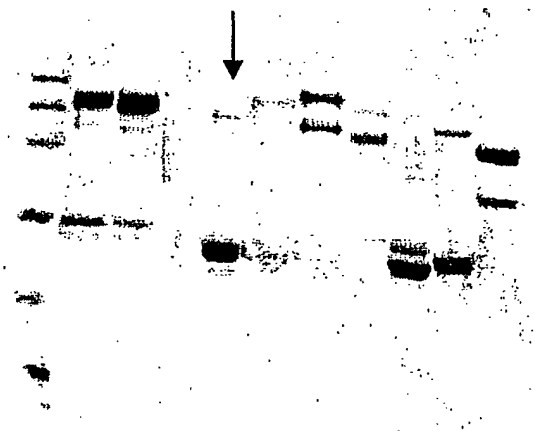




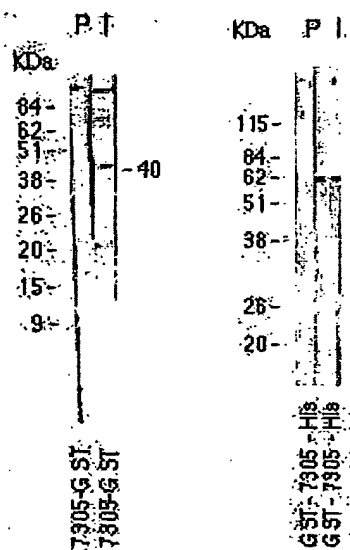
89/169

**FIGURE 89**

**Fig. 89A**



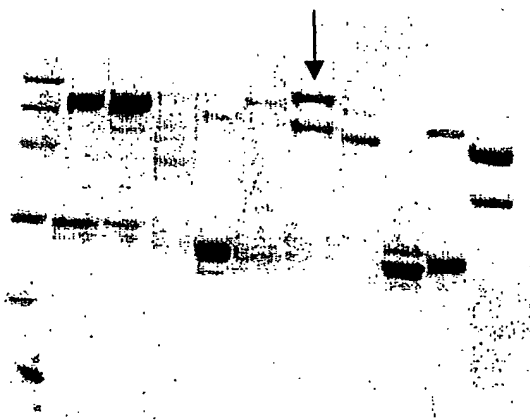
**Fig. 89B**



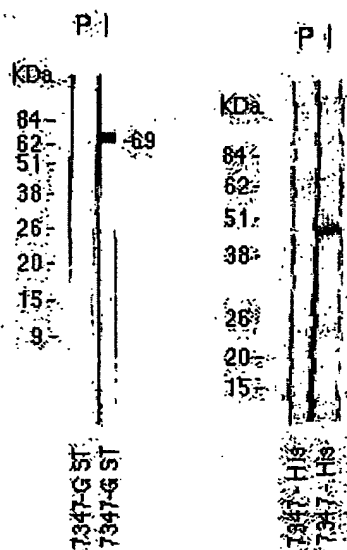
90/169

**FIGURE 90**

**Fig. 90A**



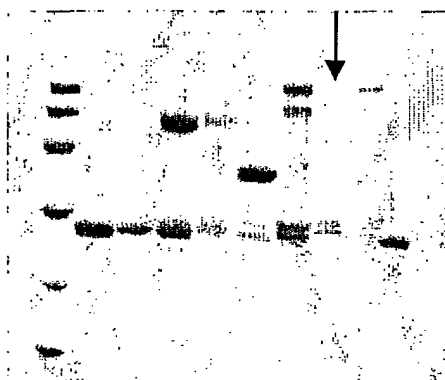
**Fig. 90B**



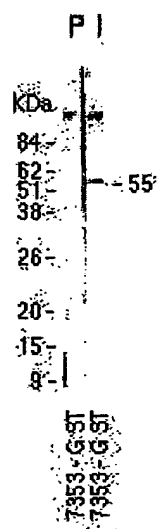
91/169

**FIGURE 91**

**Fig. 91A**



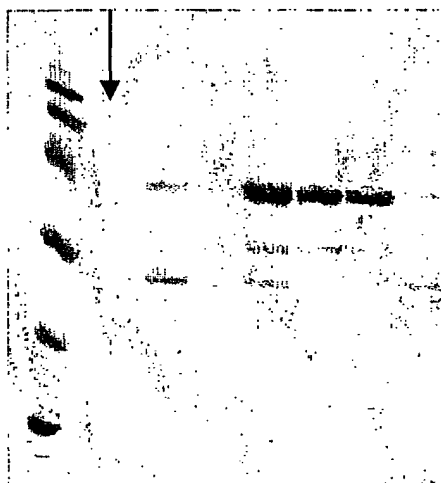
**Fig. 91B**



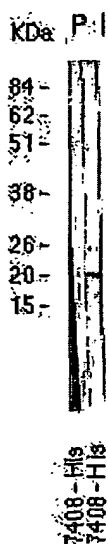
92/169

**FIGURE 92**

**FIG. 92A**



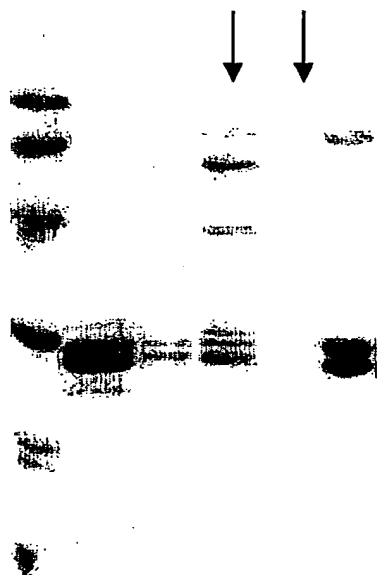
**FIG. 92B**



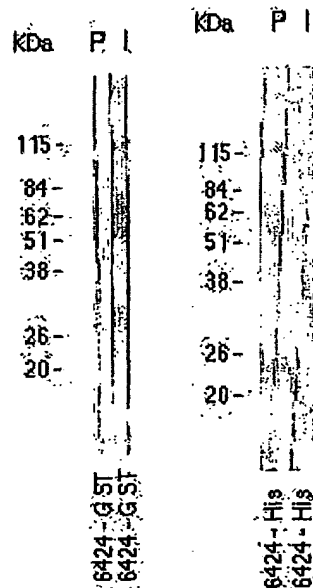
93/169

**FIGURE 93**

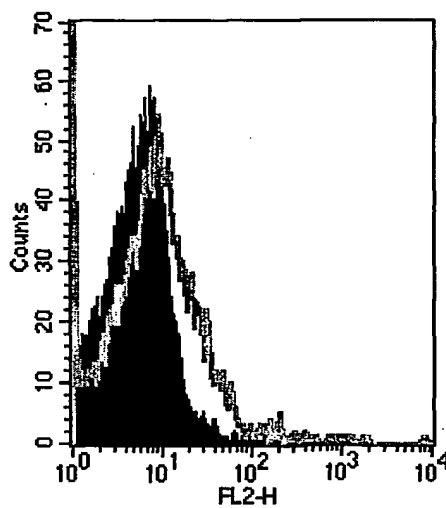
**FIG. 93A**



**FIG. 93B**



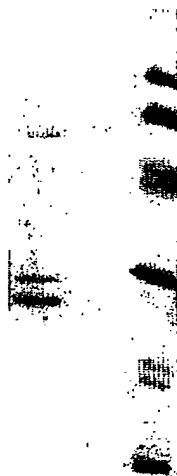
**FIG. 93C**



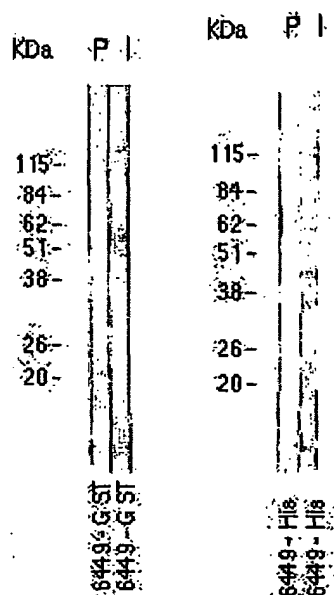
94/169

**FIGURE 94**

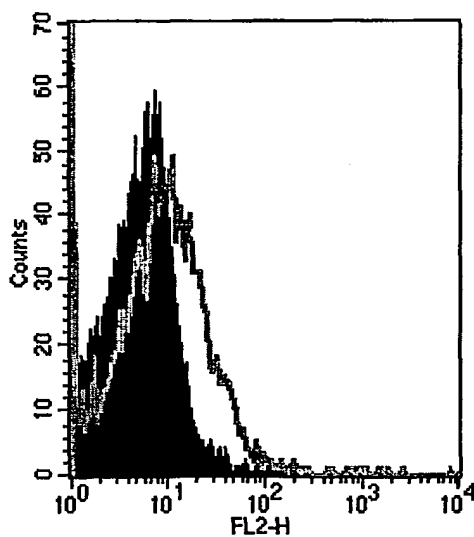
**FIG. 94A**



**FIG. 94B**



**FIG. 94C**



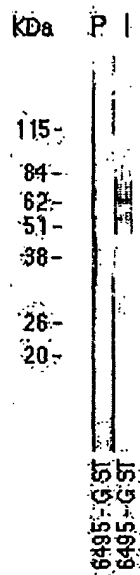
95/169

**FIGURE 95**

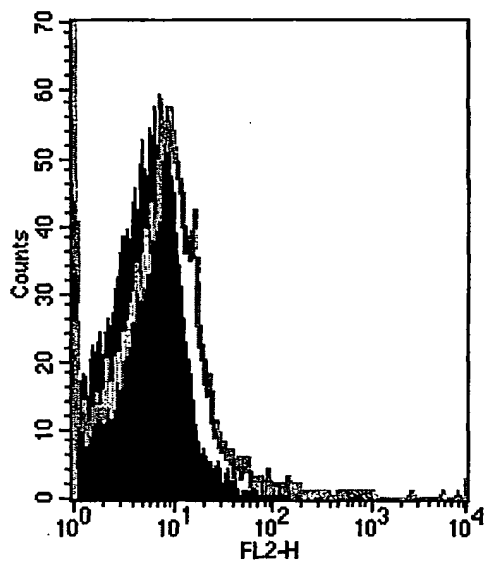
**FIG. 95A**



**FIG. 95B**



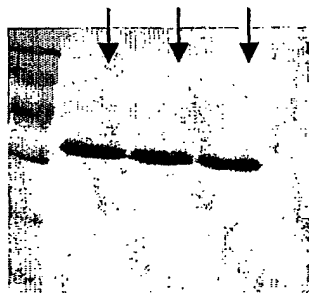
**FIG. 95C**



96/169

**FIGURE 96**

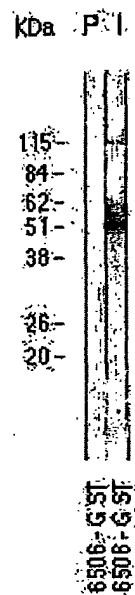
**FIG. 96A**



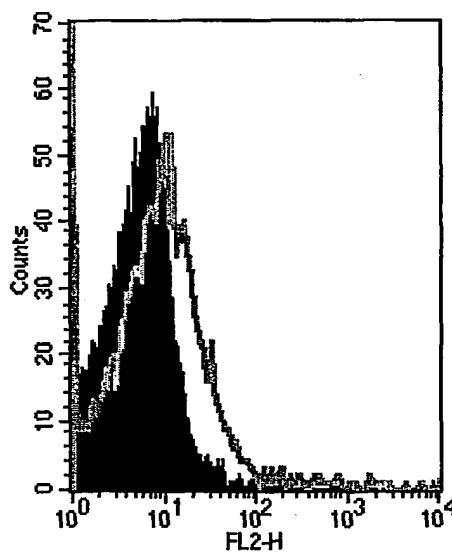
**FIG. 96B**



**FIG. 96C**



**FIG. 96D**





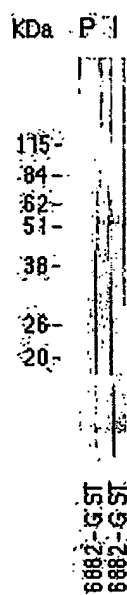
97/169

**FIGURE 97**

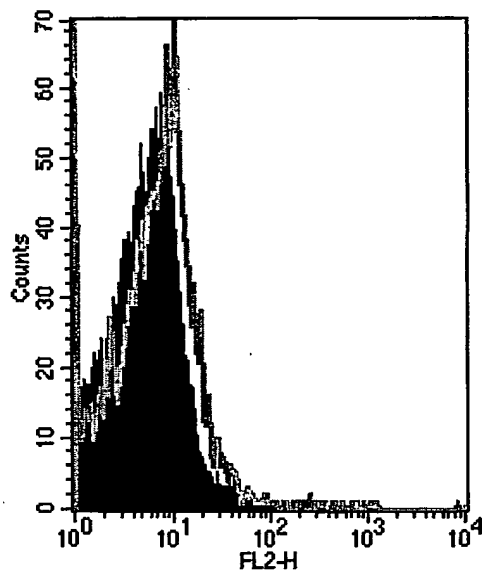
**Fig. 97A**



**Fig. 97B**



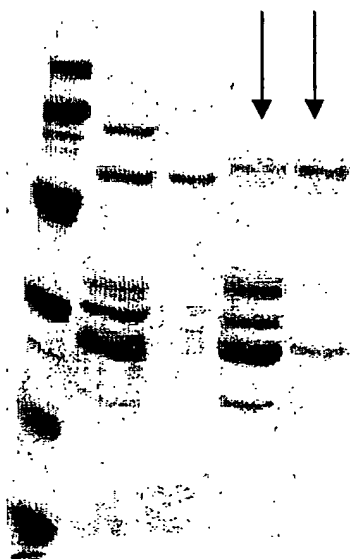
**Fig. 97C**



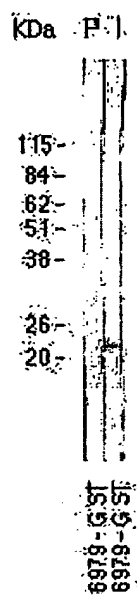
98/169

**FIGURE 98**

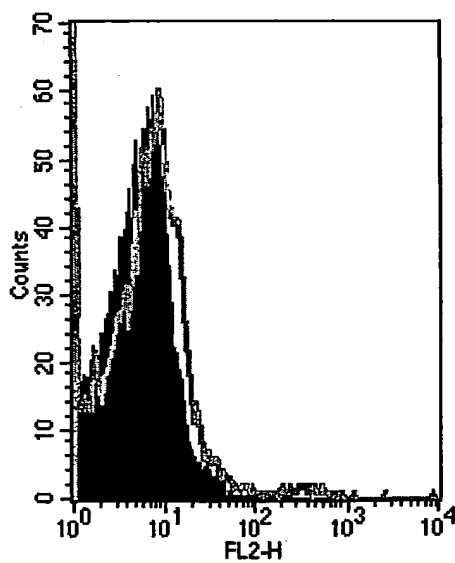
**FIG. 98A**



**FIG. 98B**



**FIG. 98C**



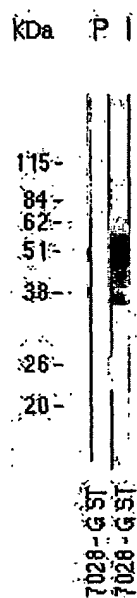
99/169

**FIGURE 99**

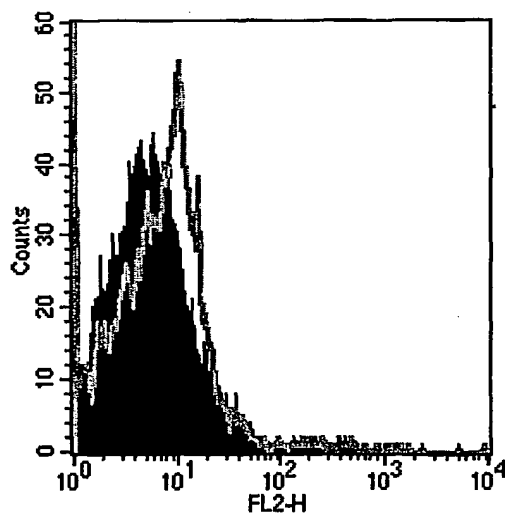
**FIG. 99A**



**FIG. 99B**



**FIG. 99C**



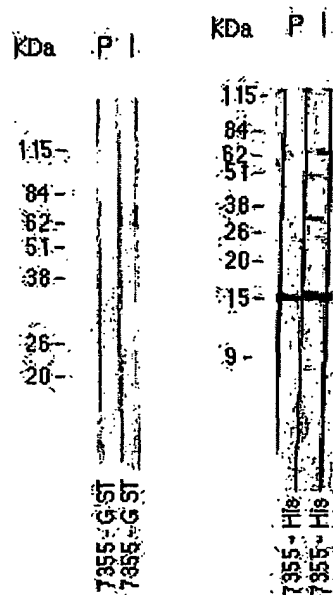
100/169

**FIGURE 100**

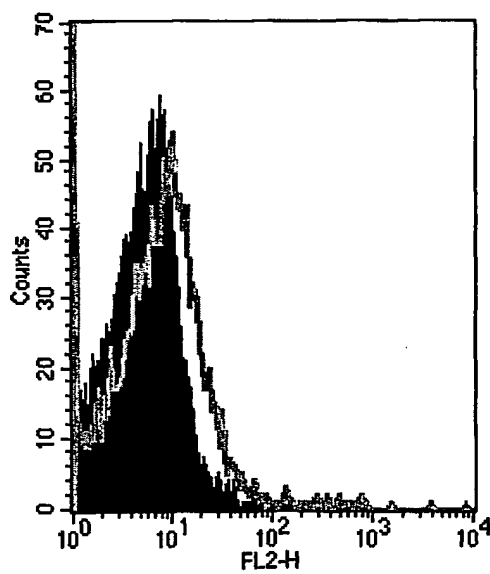
**Fig. 100A**



**Fig. 100B**



**Fig. 100C**



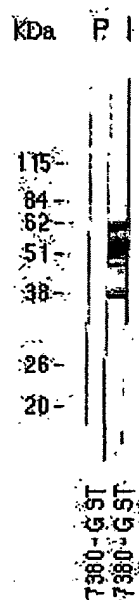
101/169

**FIGURE 101**

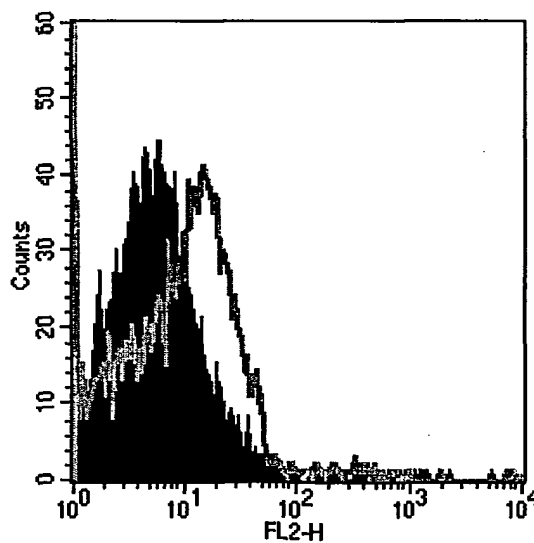
**FIG. 101A**



**FIG. 101B**



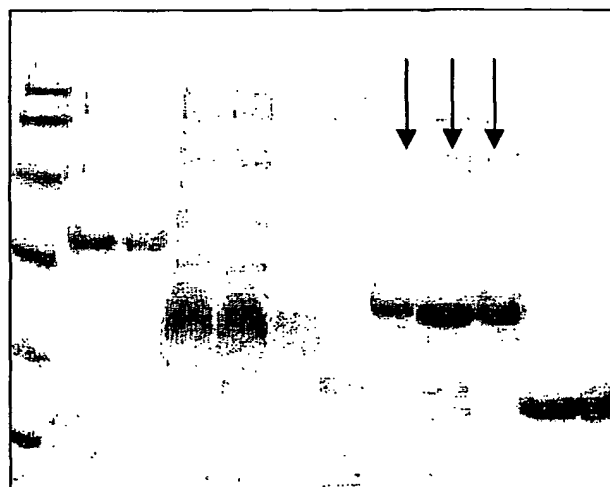
**FIG. 101C**



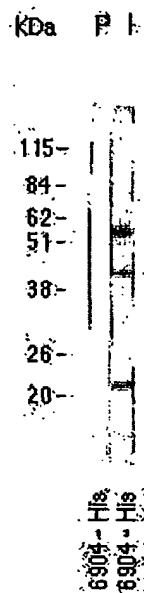
102/169

**FIGURE 102**

**Fig. 102A**

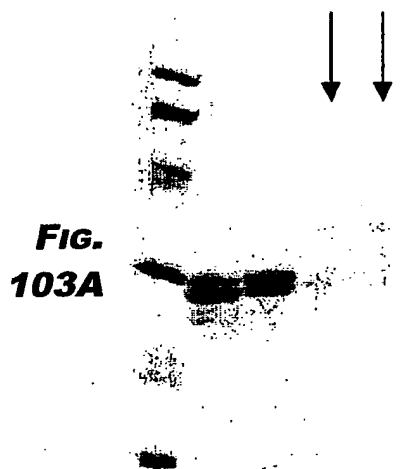


**Fig. 102B**

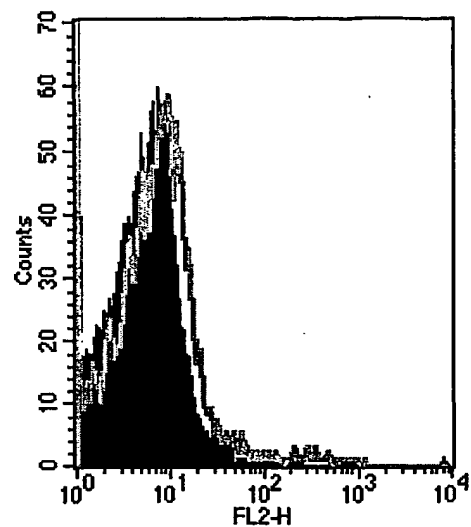


103/169

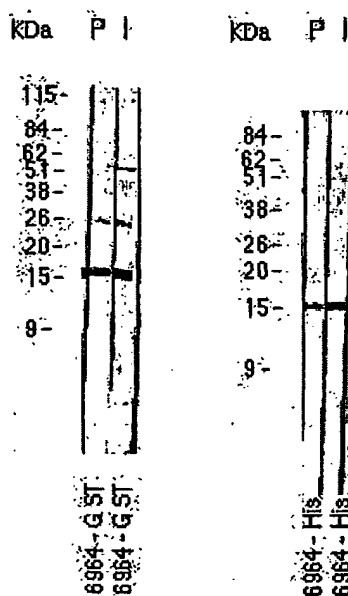
**FIGURE 103**



**FIG. 103C**



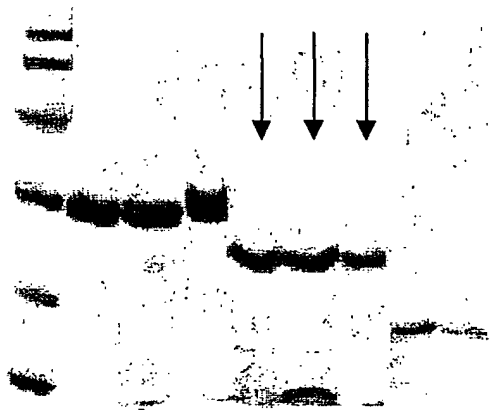
**FIG. 103B**



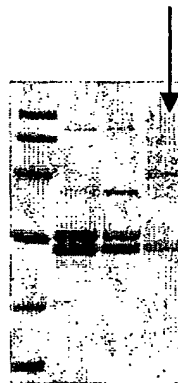
104/169

**FIGURE 104**

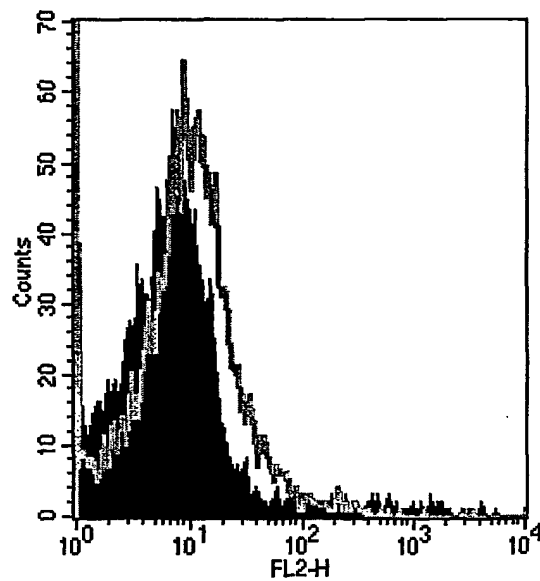
**Fig. 104A**



**FIG. 104B**



**Fig. 104C**

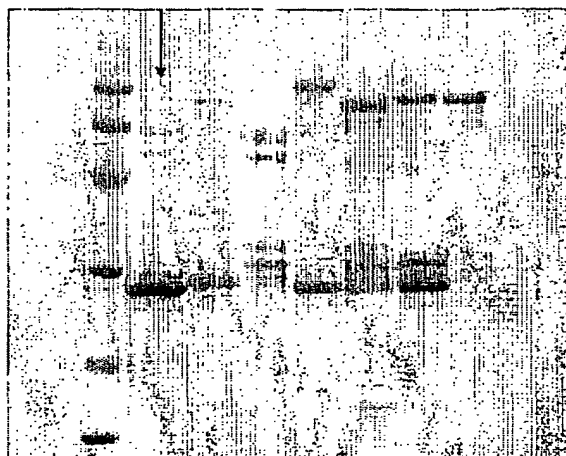




105/169

**FIGURE 105**

**Fig. 105A**



kDa P I

**Fig. 105B**

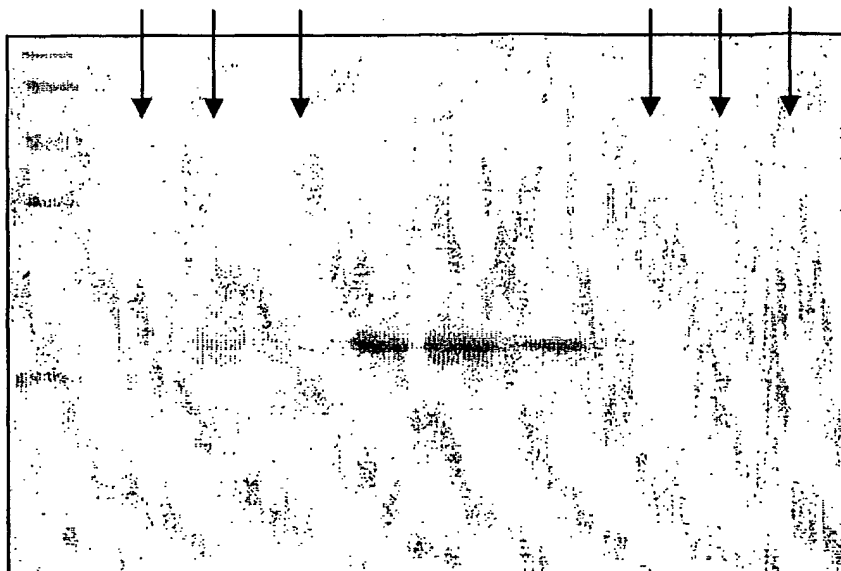
115-  
84-  
62-  
51-  
38-  
26-  
20-

6201-G-5T  
6201-G-5T

106/169

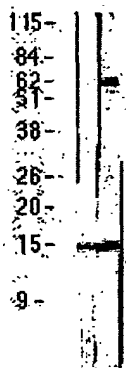
**FIGURE 106**

**Fig. 106A**



**FIG. 106B**

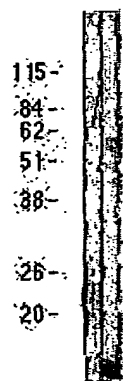
kDa P I.



His His  
6306 6306

**FIGURE 107**

kDa P I.



His His  
6434 6434

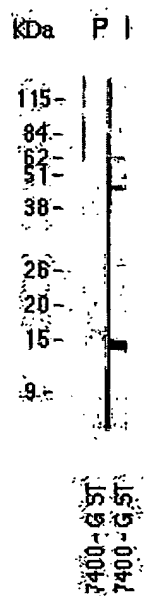
107/169

**FIGURE 108**

**Fig. 108A**



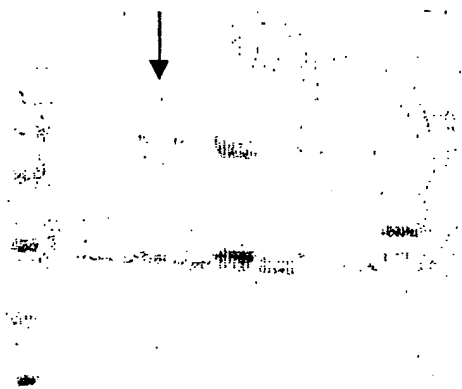
**Fig. 108B**



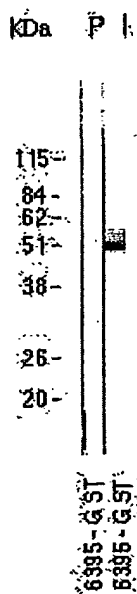
108/169

**FIGURE 109**

**Fig. 109A**

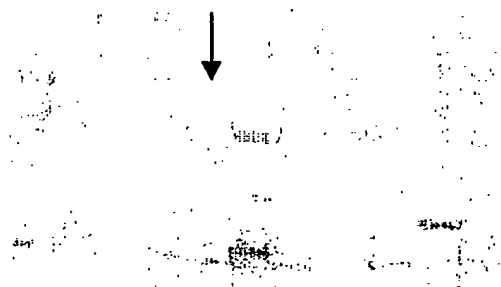


**Fig. 109B**

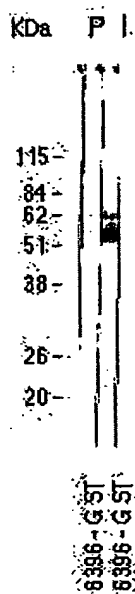


109/169

**FIGURE 110**



**Fig. 110A**

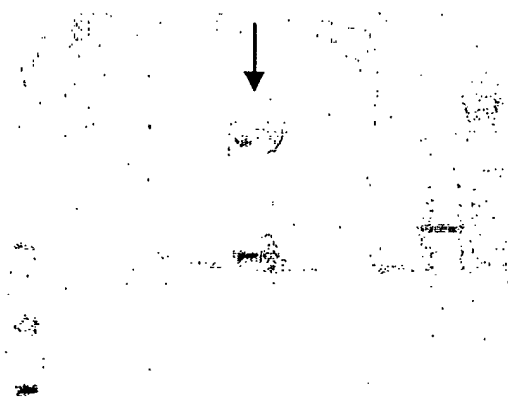


**Fig. 110B**

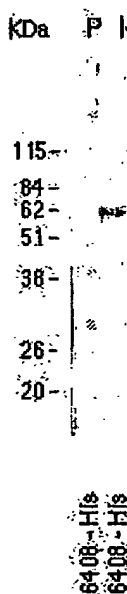
110/169

**FIGURE 111**

**FIG. 111A**

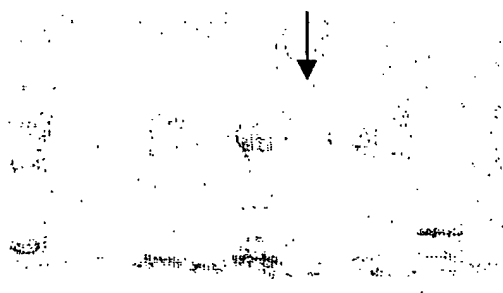


**FIG. 111B**

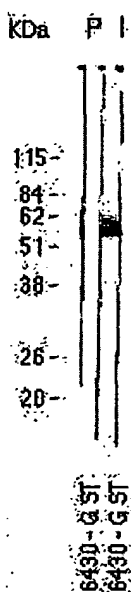


111/169

**FIGURE 112**



**FIG. 112A**



**FIG. 112B**

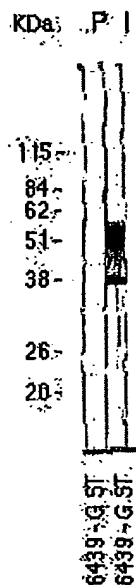
112/169

**FIGURE 113**

**Fig. 113A**

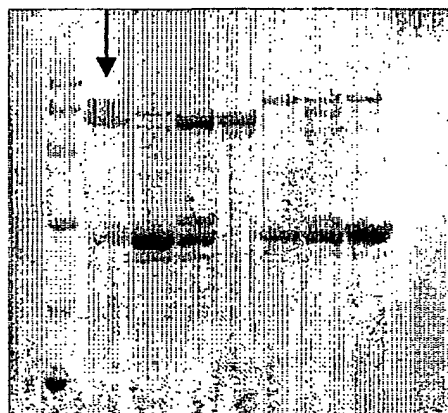
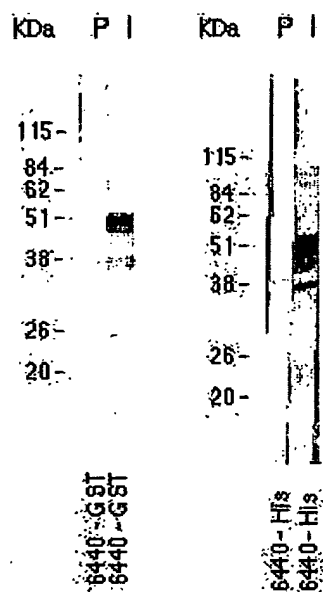


**Fig. 113B**





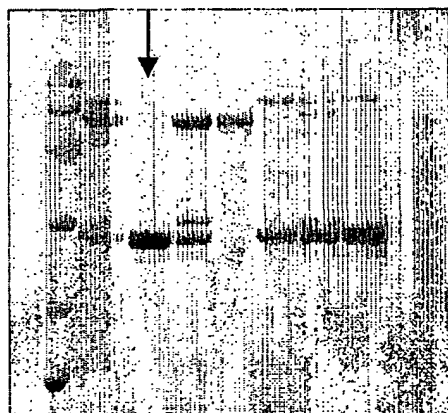
113/169

**FIGURE 114****FIG. 114A****FIG. 114B**

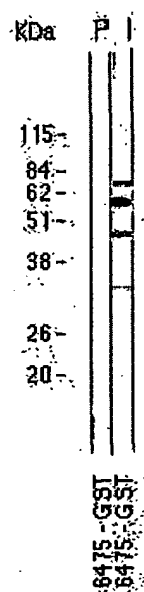
114/169

**FIGURE 115**

**Fig. 115A**



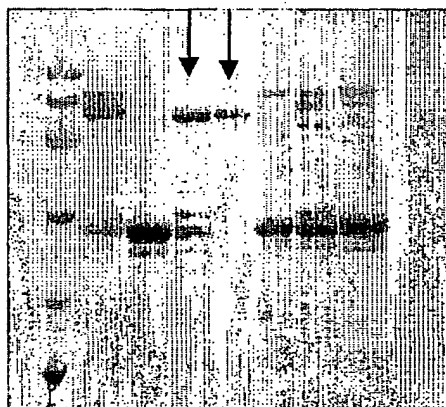
**Fig. 115B**



115/169

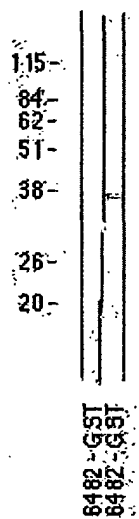
**FIGURE 116**

**FIG. 116A**



kDa P I

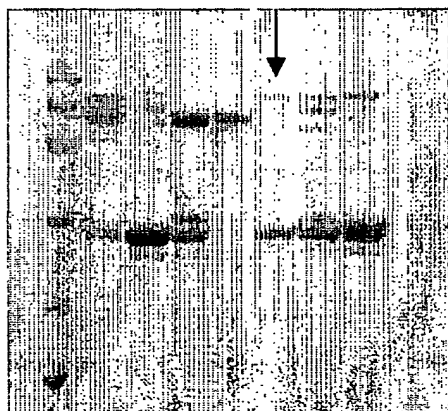
**FIG. 116B**



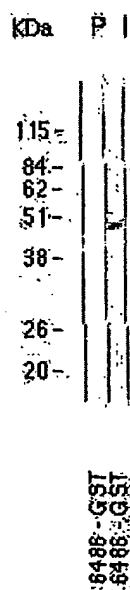
116/169

**FIGURE 117**

**Fig. 117A**



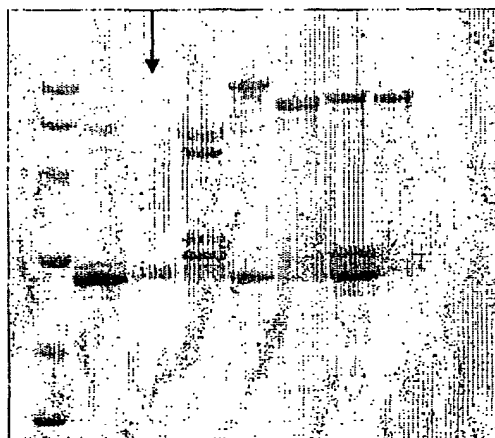
**Fig. 117B**



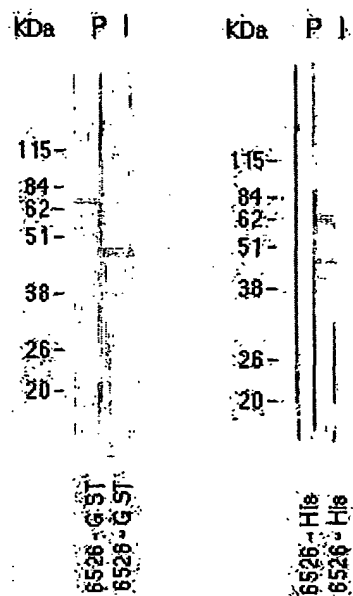
117/169

**FIGURE 118**

**Fig. 118A**



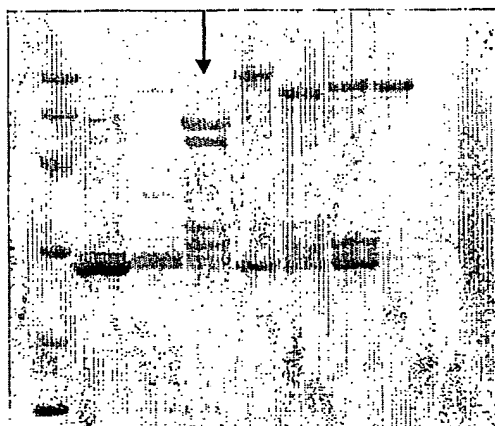
**Fig. 118B**



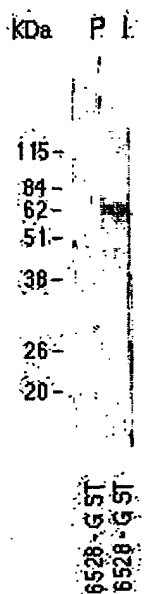
118/169

**FIGURE 119**

**Fig. 119A**



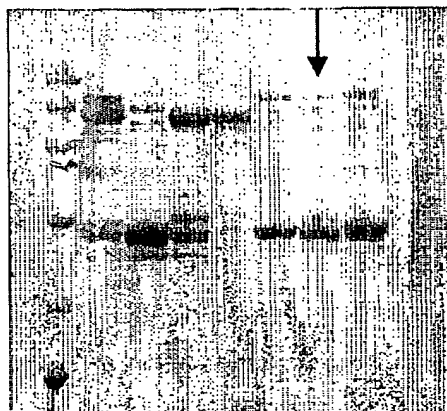
**Fig. 119B**



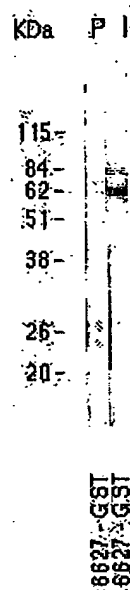
119/169

**FIGURE 120**

**FIG. 120A**



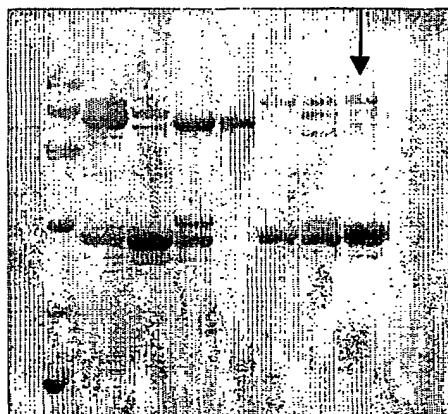
**FIG. 120B**



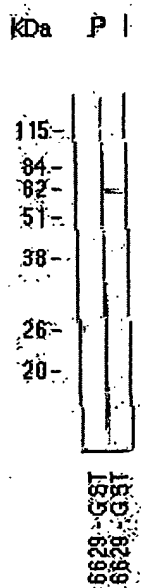
120/169

**FIGURE 121**

**FIG. 121A**



**FIG. 121B**

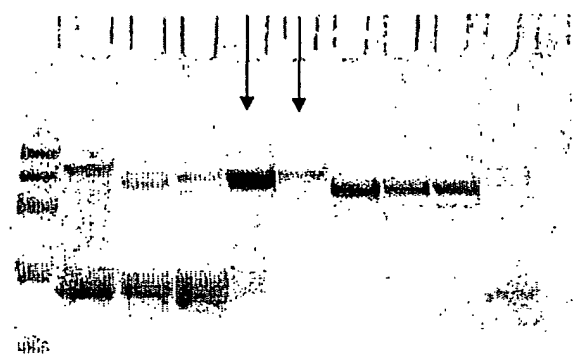




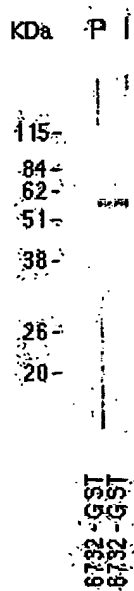
121/169

**FIGURE 122**

**Fig. 122A**



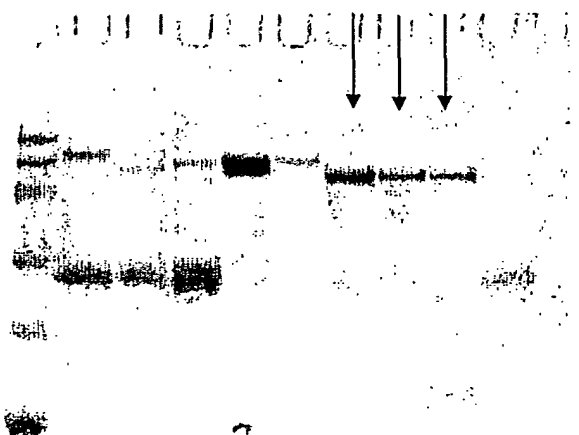
**Fig. 122B**



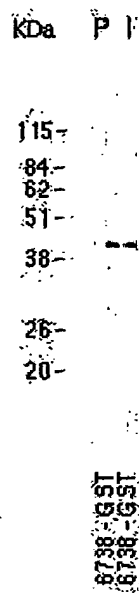
122/169

**FIGURE 123**

**Fig. 123A**



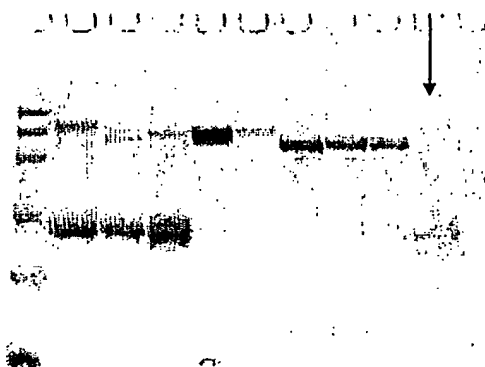
**Fig. 123B**



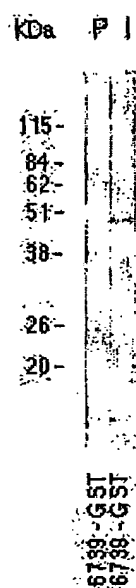
123/169

**FIGURE 124**

**Fig. 124A**



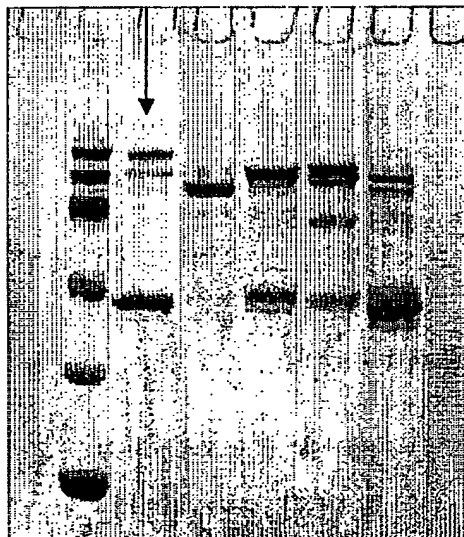
**Fig. 124B**



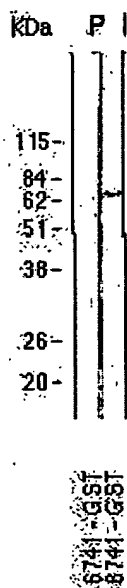
124/169

**FIGURE 125**

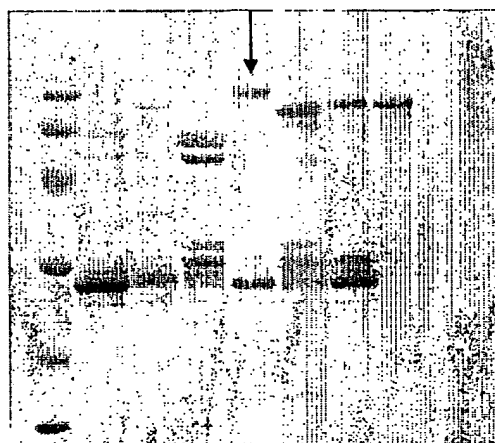
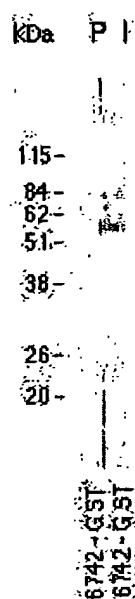
**Fig. 125A**



**Fig. 125B**

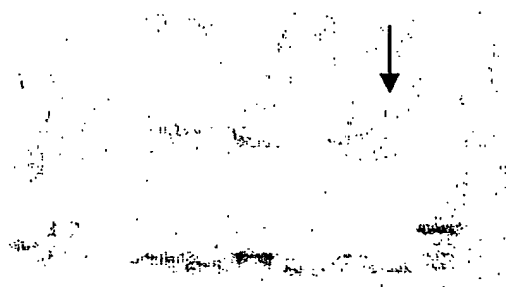


125/169

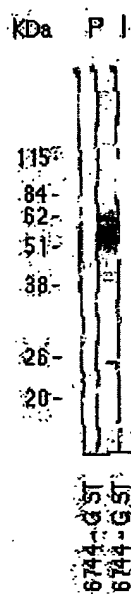
**FIGURE 126****FIG. 126A****FIG. 126B**

126/169

**FIGURE 127**



**Fig. 127A**

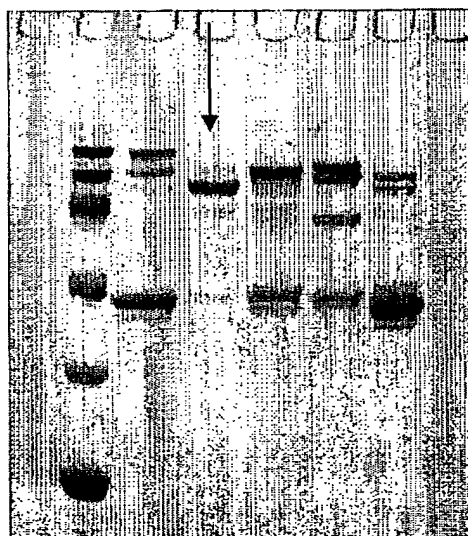


**Fig. 127B**

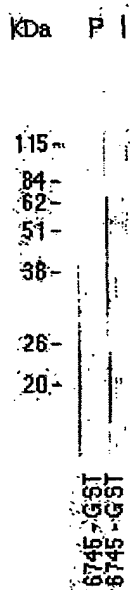
127/169

**FIGURE 128**

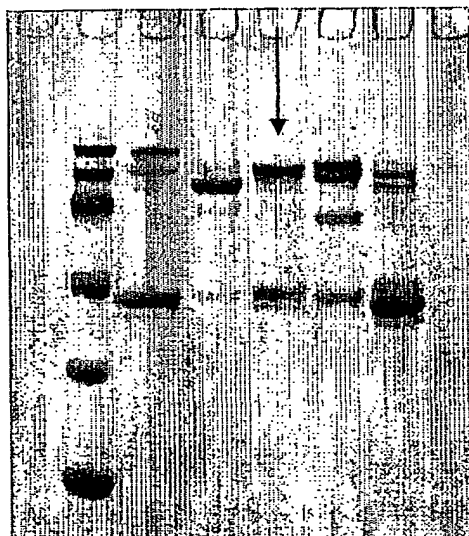
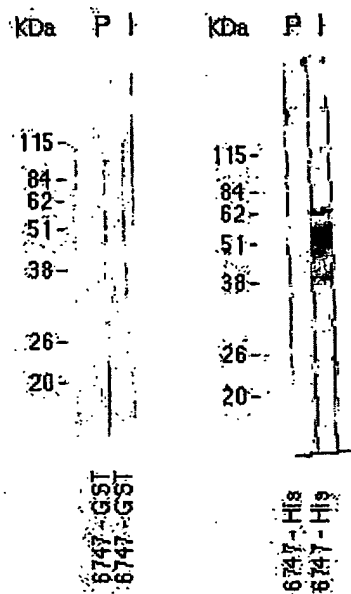
**Fig. 128A**



**Fig. 128B**



128/169

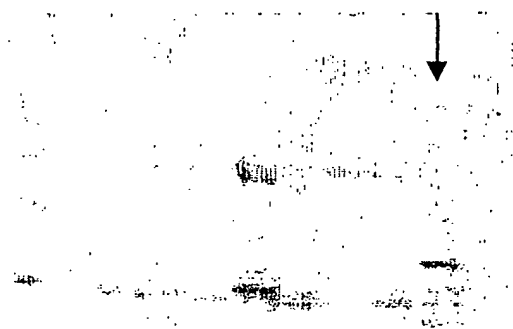
**FIGURE 129****Fig. 129A****Fig. 129B**



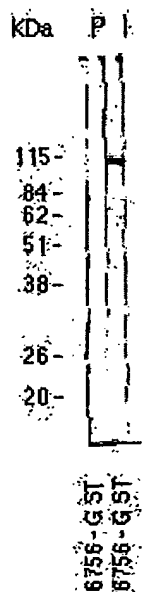
129/169

**FIGURE 130**

**FIG. 130A**



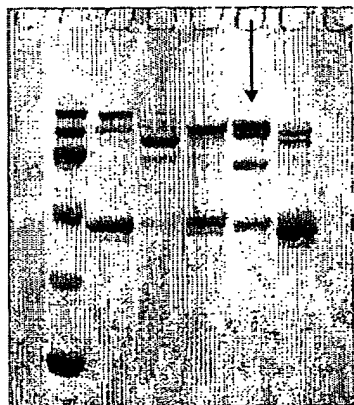
**FIG. 130B**



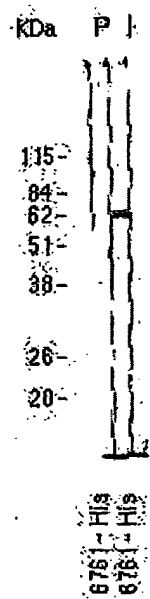
130/169

**FIGURE 131**

**Fig. 131A**



**Fig. 131B**

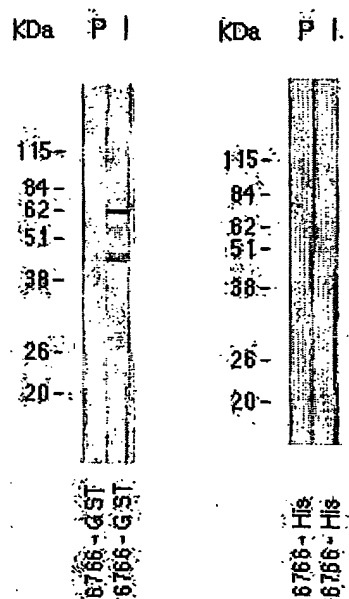


131/169

**FIGURE 132**



**Fig. 132A**

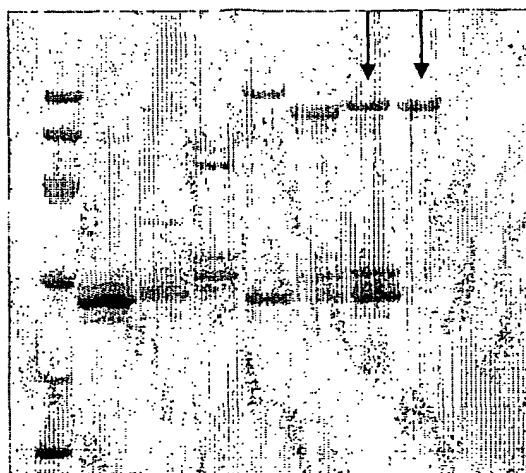


**Fig. 132B**

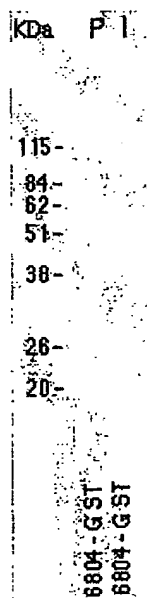
132/169

**FIGURE 133**

**Fig. 133A**



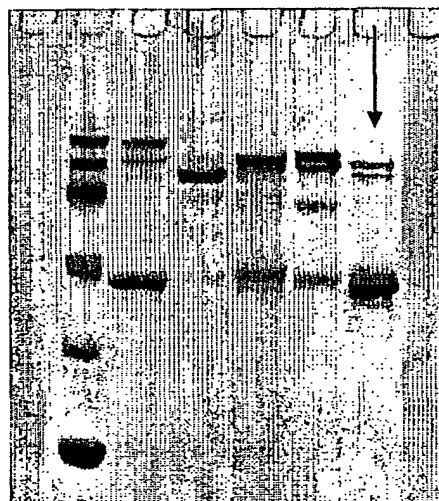
**Fig. 133B**



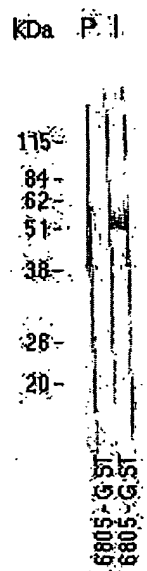
133/169

**FIGURE 134**

**FIG. 134A**



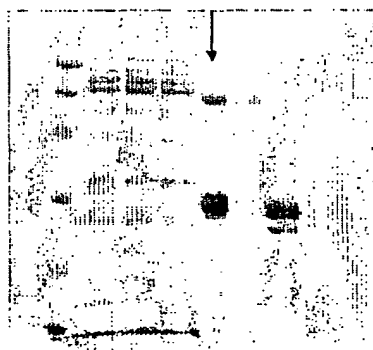
**FIG. 134B**



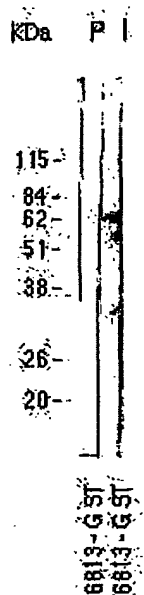
134/169

**FIGURE 135**

**FIG. 135A**



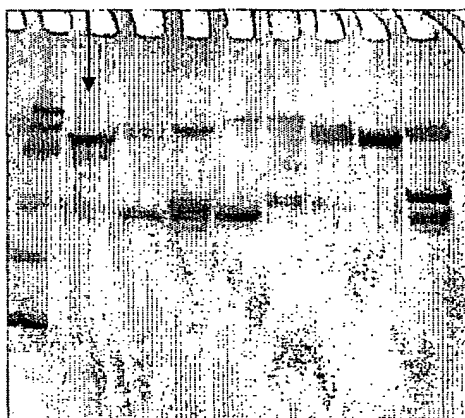
**FIG. 135B**



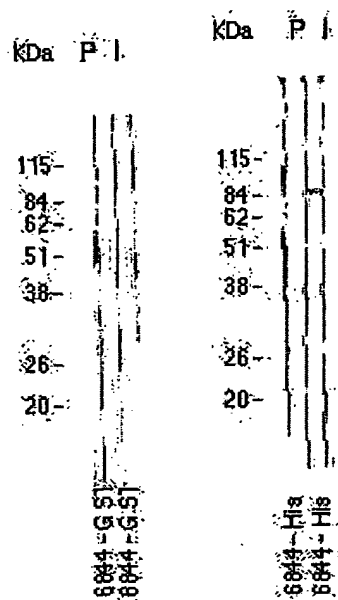
135/169

**FIGURE 136**

**Fig. 136A**



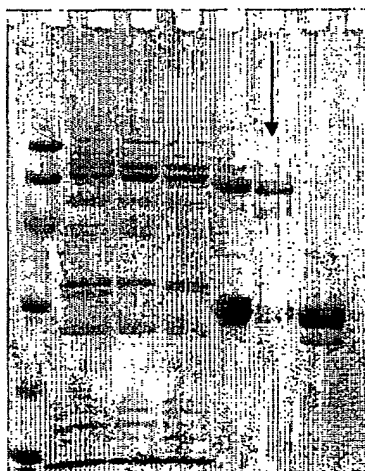
**Fig. 136B**



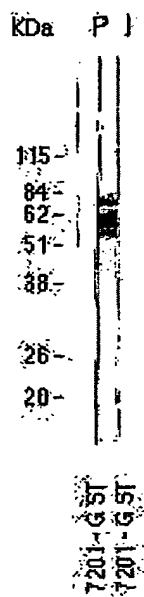
136/169

**FIGURE 137**

**FIG. 137A**



**FIG. 137B**

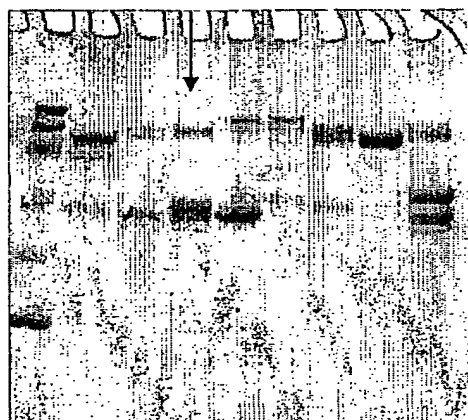




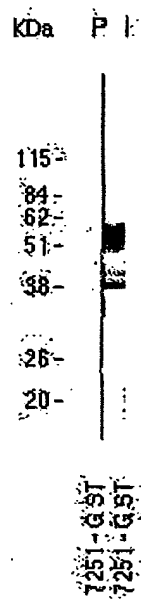
137/169

**FIGURE 138**

**FIG. 138A**



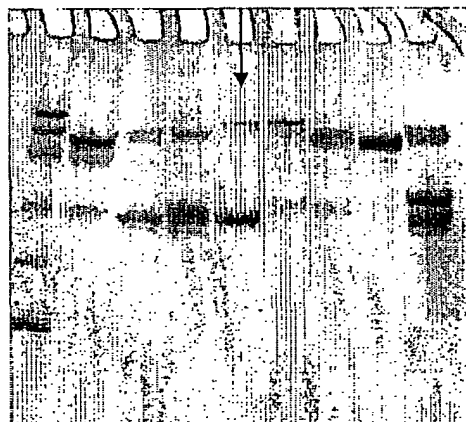
**FIG. 138B**



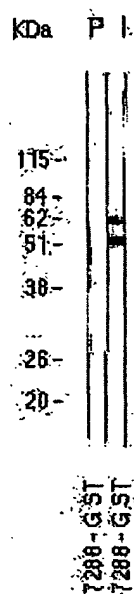
138/169

**FIGURE 139**

**Fig. 139A**



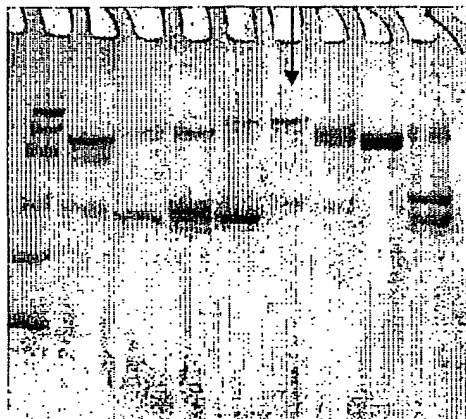
**Fig. 139B**



139/169

**FIGURE 140**

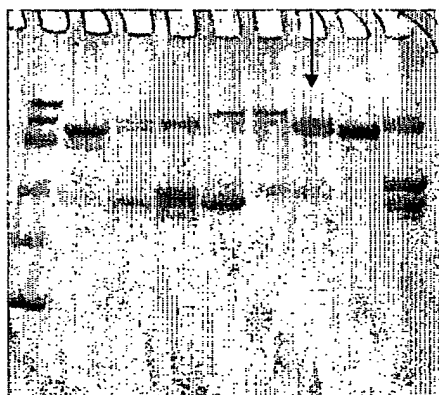
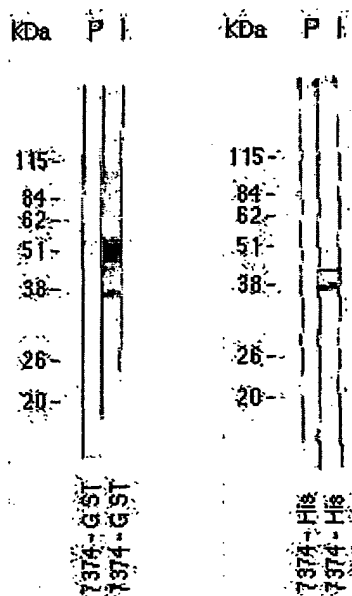
**Fig. 140A**



**Fig. 140B**

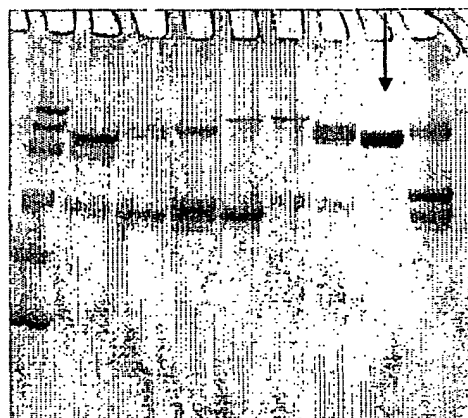


140/169

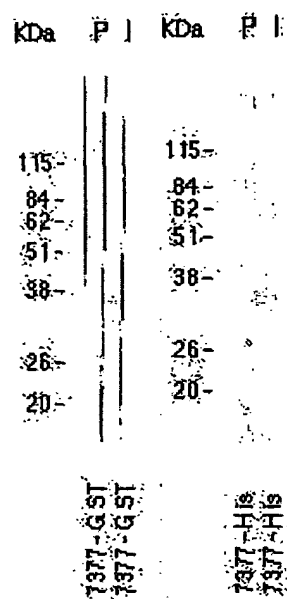
**FIGURE 141****Fig. 141A****Fig. 141B**

141/169

**FIGURE 142**



**FIG. 142A**

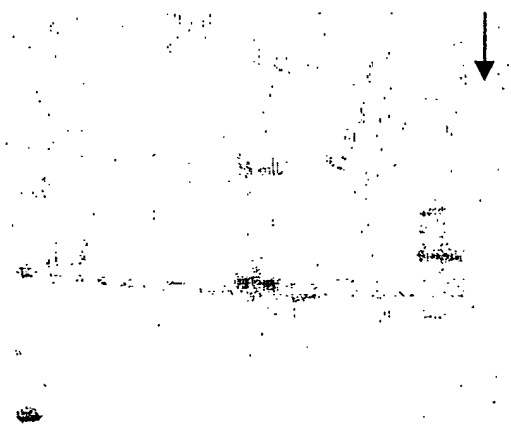


**FIG. 142B**

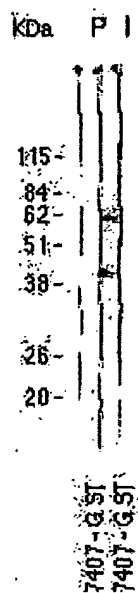
142/169

**FIGURE 143**

**Fig. 143A**



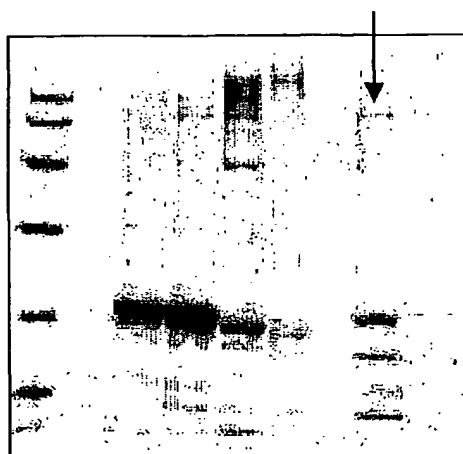
**Fig. 143B**



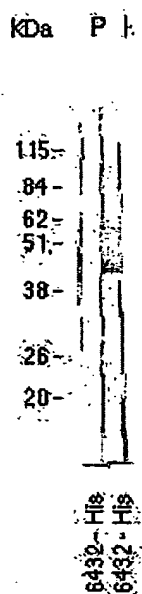
143/169

**FIGURE 144**

**Fig. 144A**



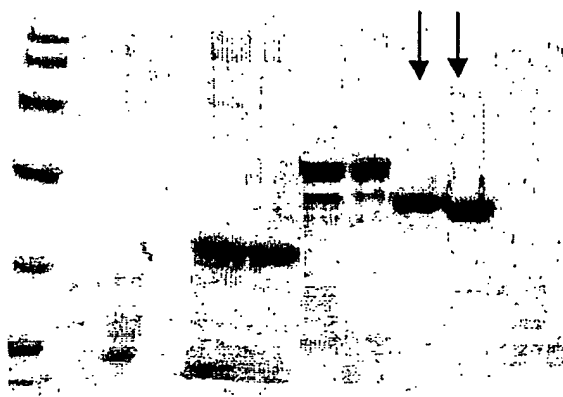
**Fig. 144B**



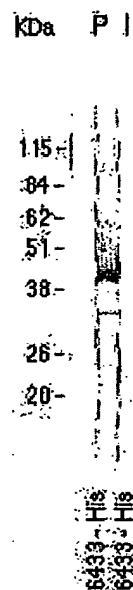
144/169

**FIGURE 145**

**Fig. 145A**



**Fig. 145B**

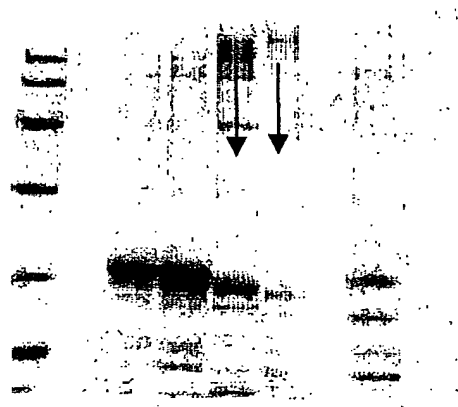




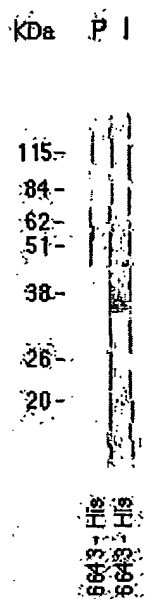
145/169

**FIGURE 146**

**FIG. 146A**



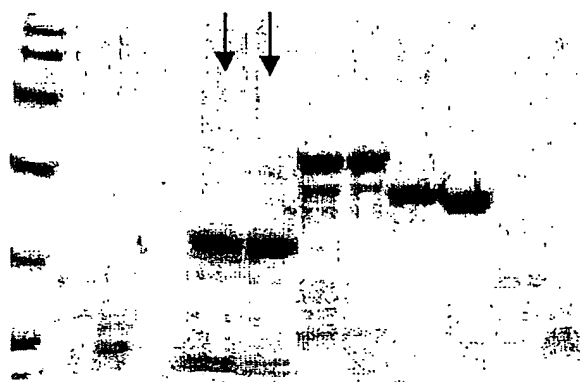
**FIG. 146B**



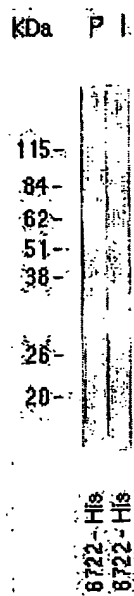
146/169

**FIGURE 147**

**FIG. 147A**



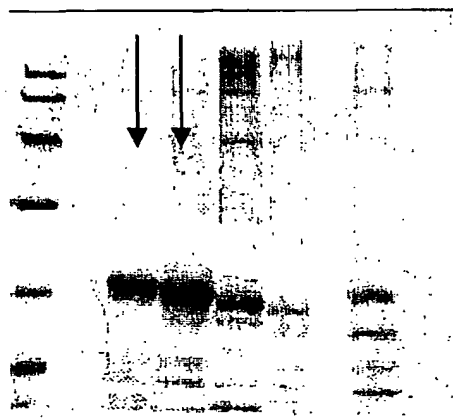
**FIG. 147B**



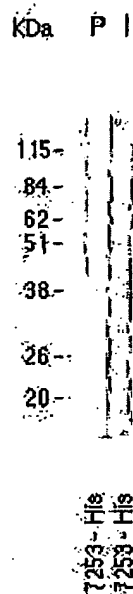
147/169

**FIGURE 148**

**Fig. 148A**



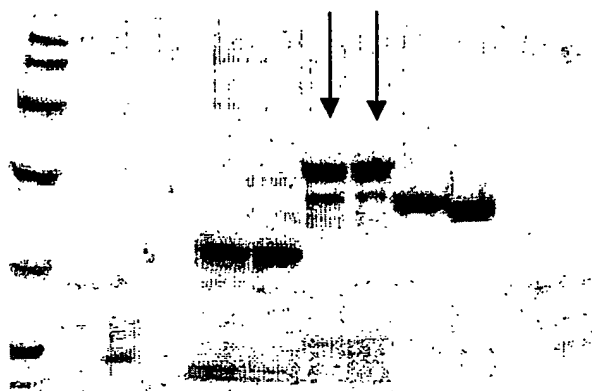
**Fig. 148B**



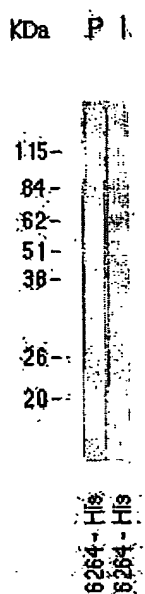
148/169

**FIGURE 149**

**Fig. 149A**



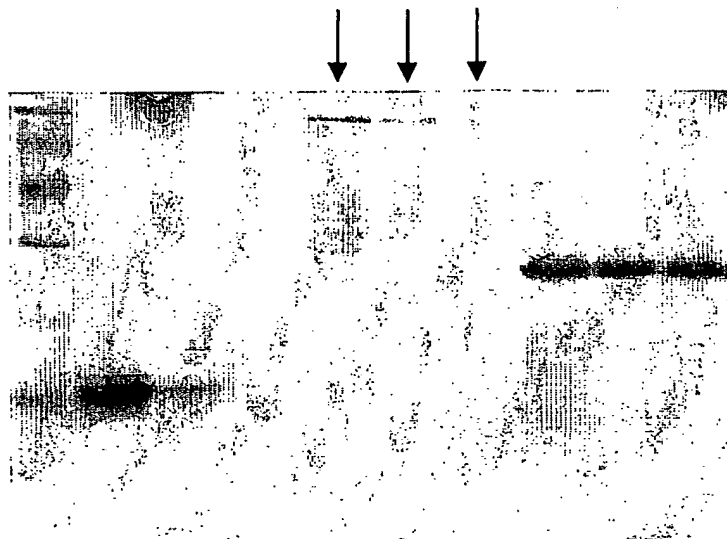
**Fig. 149B**



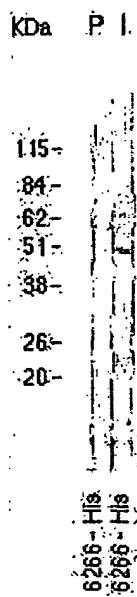
149/169

**FIGURE 150**

**Fig. 150A**

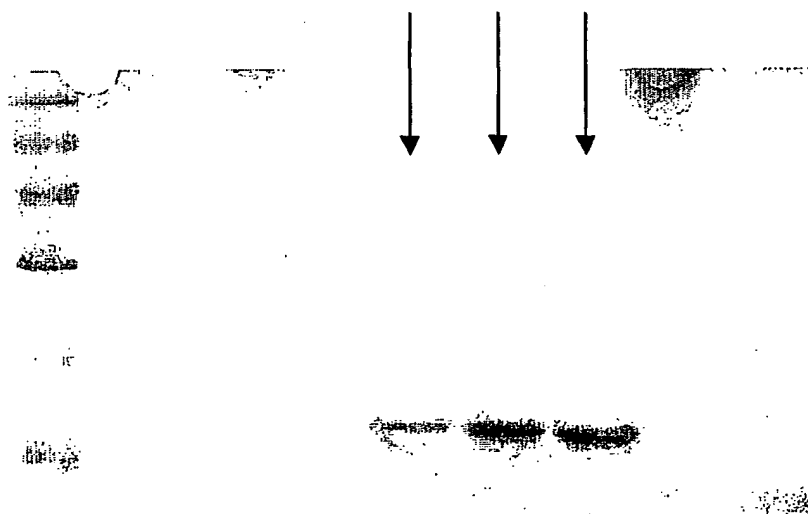


**Fig. 150B**



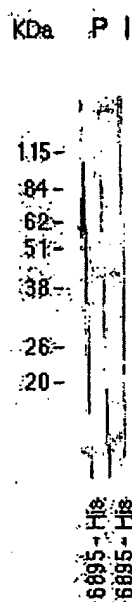
150/169

**FIGURE 151**



**Fig. 151A**

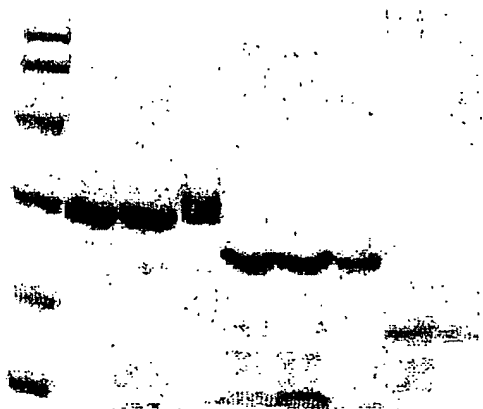
**Fig. 151B**



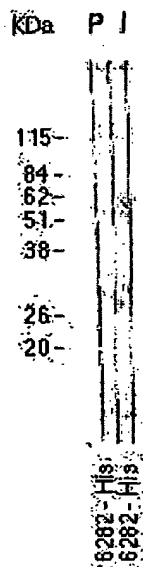
151/169

**FIGURE 152**

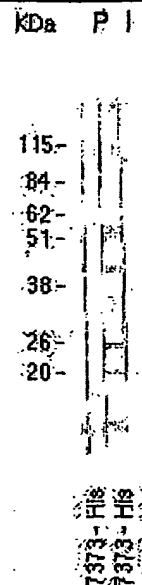
**Fig. 152A**



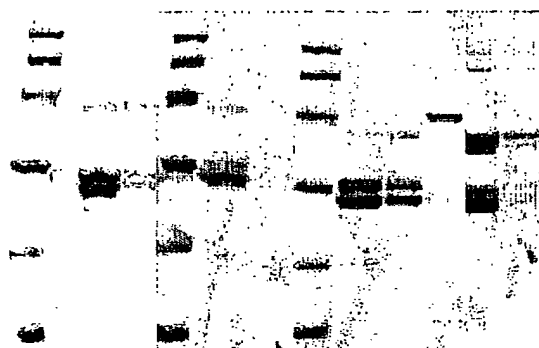
**Fig. 152B**



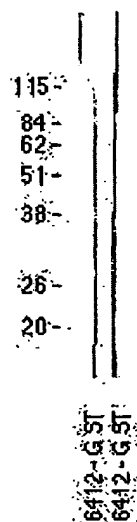
**FIGURE 153**



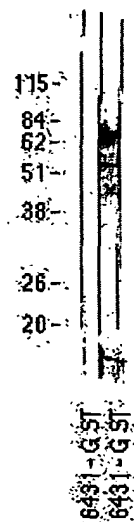
152/169

**FIGURE 154****Fig. 154A****FIG. 154B**

kDa P I

**FIGURE 155**

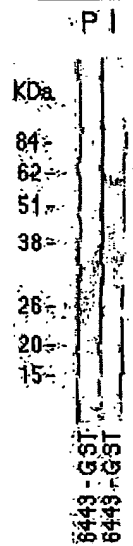
kDa P I



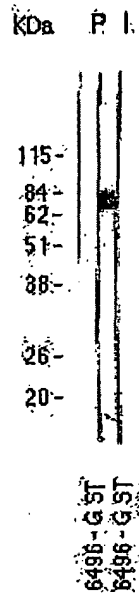


153/169

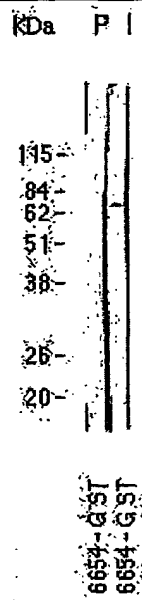
**FIGURE 156**



**FIGURE 157**



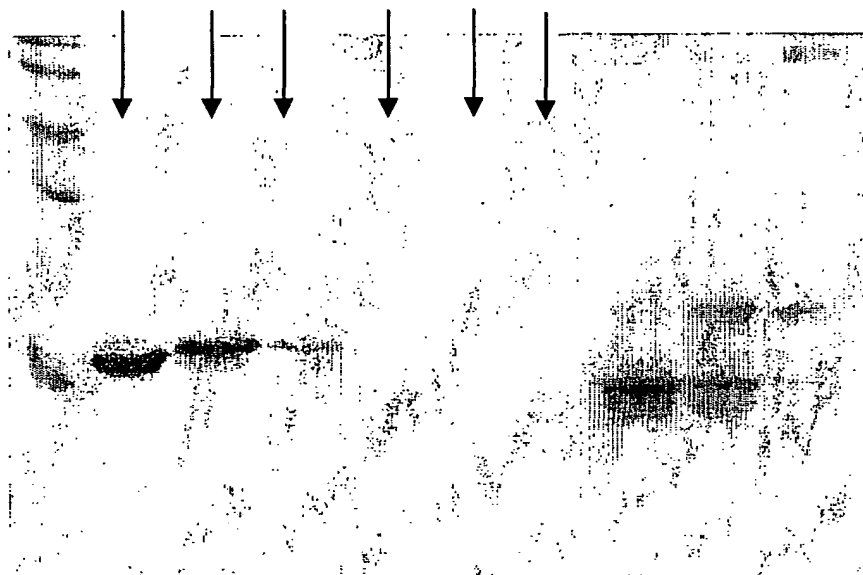
**FIGURE 158**



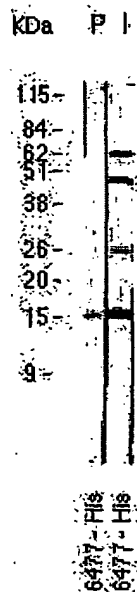
154/169

**FIGURE 159**

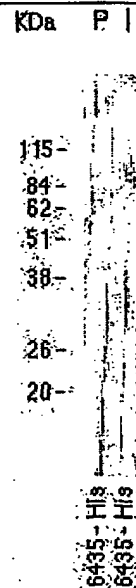
**Fig. 159A**



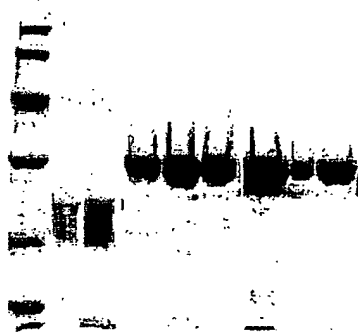
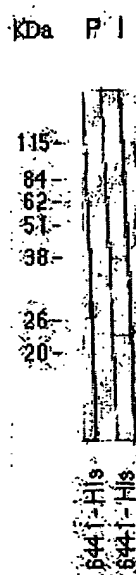
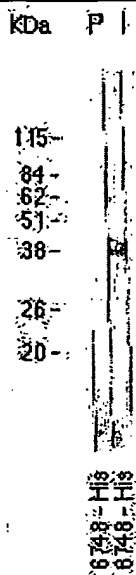
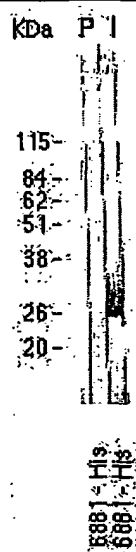
**Fig. 159B**



**FIGURE 160**



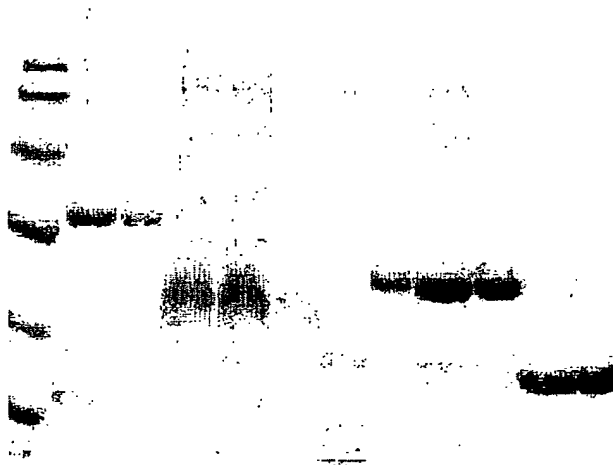
155/169

**FIGURE 161****Fig. 161A****FIG. 161B****FIGURE 162****FIGURE 163**

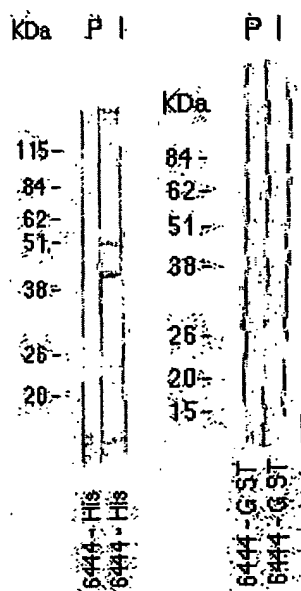
156/169

**FIGURE 164**

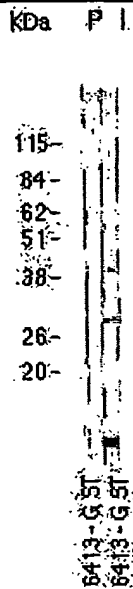
**Fig. 164A**



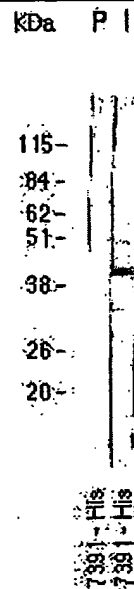
**Fig. 164B**



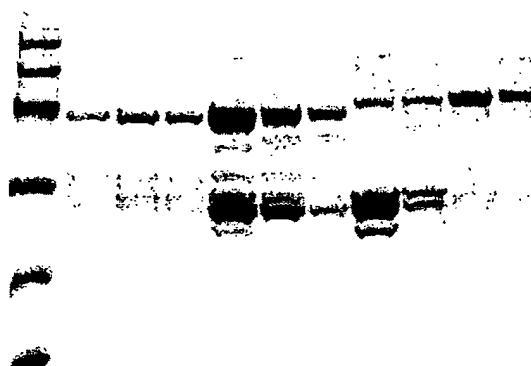
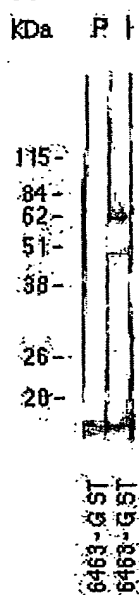
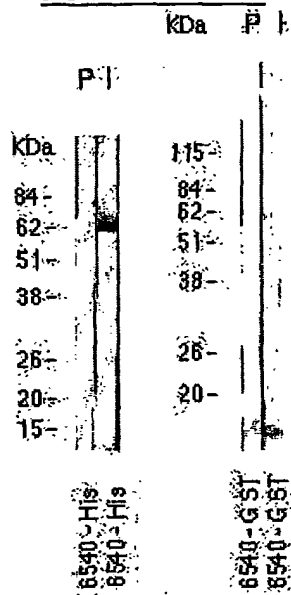
**FIGURE 165**



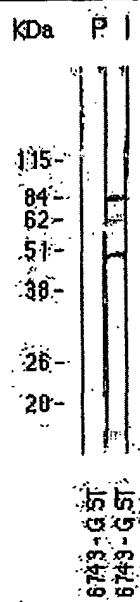
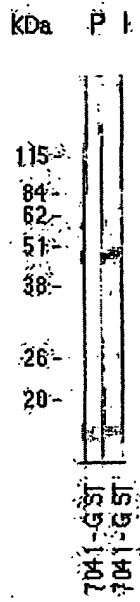
**FIGURE 166**



157/169

**FIGURE 167****FIG. 167A****FIG. 167B****FIGURE 168**

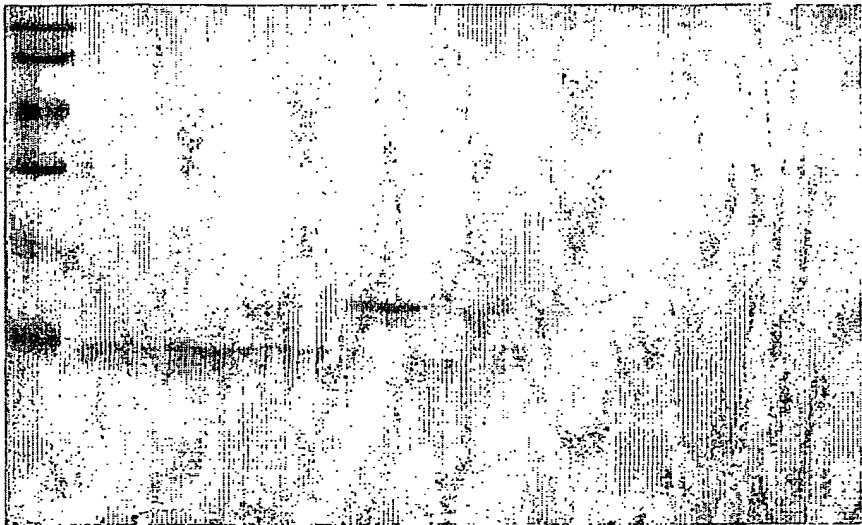
158/169

**FIGURE 169****FIGURE 170**

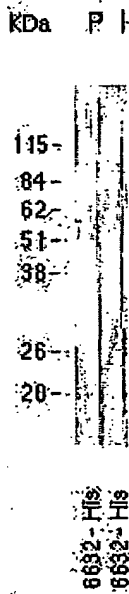
159/169

**FIGURE 171**

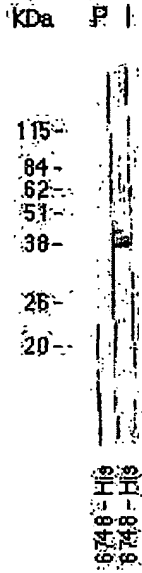
**FIG. 171A**



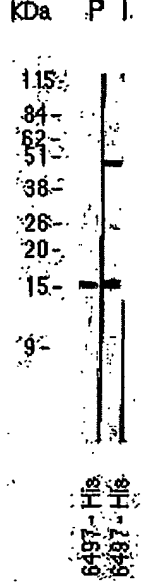
**FIG. 171B**



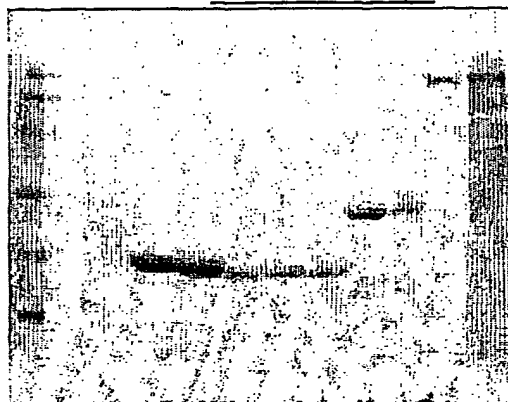
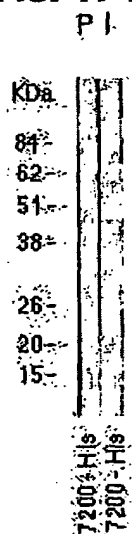
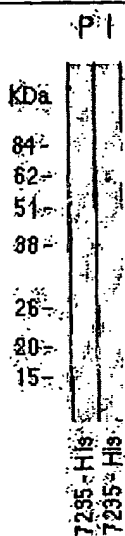
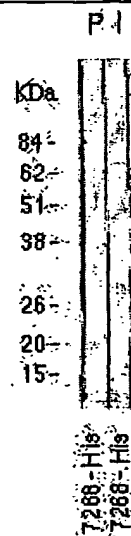
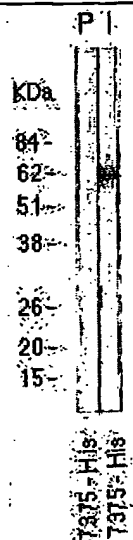
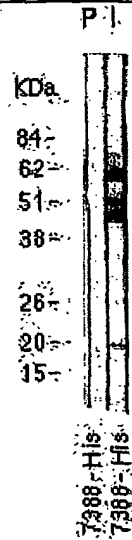
**FIGURE 172**



**FIGURE 173**



160/169

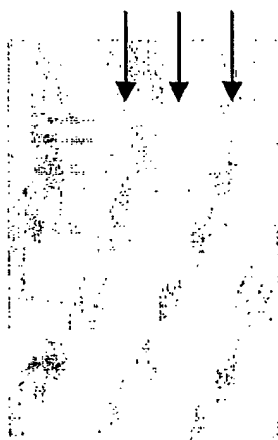
**FIGURE 174****Fig. 174A****FIG. 174B****FIGURE 175****FIGURE 176****FIGURE 177****FIGURE 178**



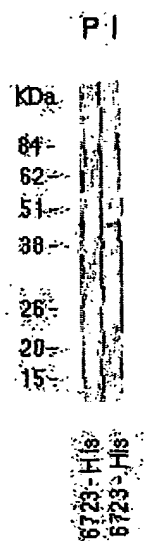
161/169

**FIGURE 179**

**Fig. 179A**



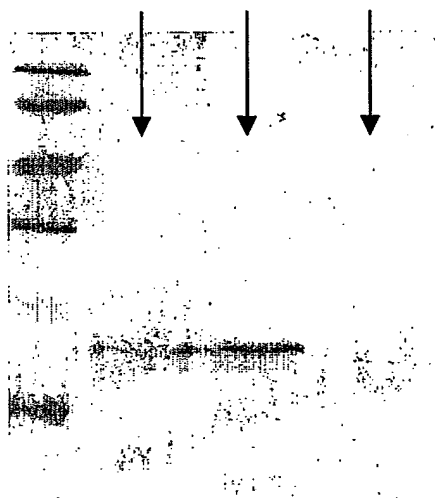
**Fig. 179B**



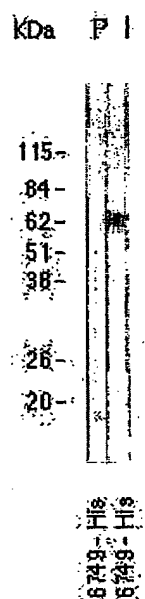
162/169

**FIGURE 180**

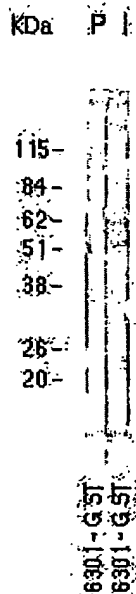
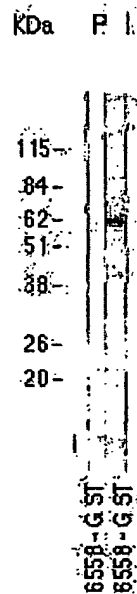
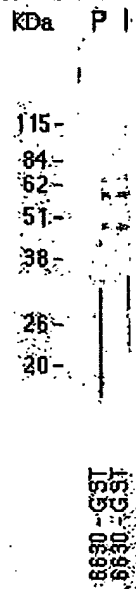
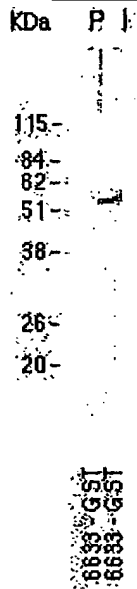
**Fig. 180A**



**Fig. 180B**



163/169

**FIGURE 181****FIGURE 182****FIGURE 183****FIGURE 184**

164/169

**FIGURE 185**

KDa P I

115-

84-

62-

51-

38-

26-

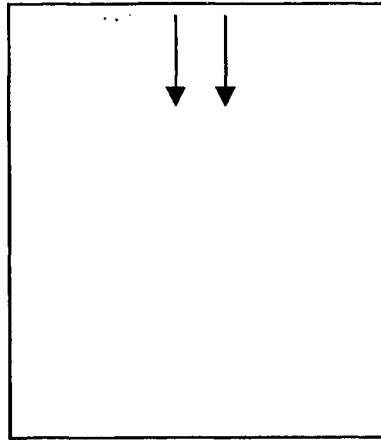
20-

66k2-GST  
66k2-GST

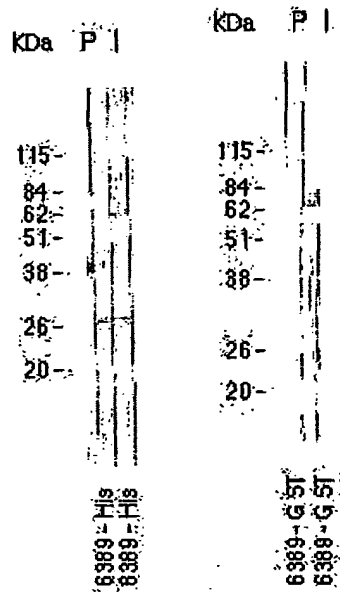
165/169

**FIGURE 186**

**FIG. 186A**



**FIG. 186B**



166/169

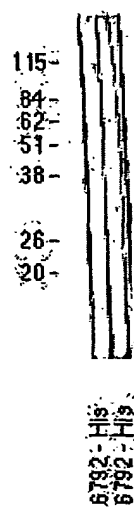
**FIGURE 187**

**Fig. 187A**



KDa P I

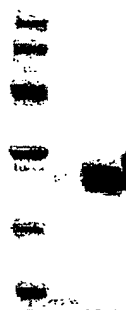
**Fig. 187B**



167/169

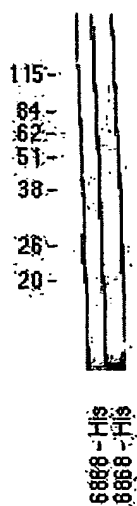
**FIGURE 188**

**FIG. 188A**



KDa P. I

**FIG. 188B**



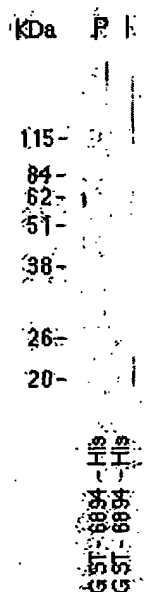
168/169

**FIGURE 189**

**FIG. 189A**

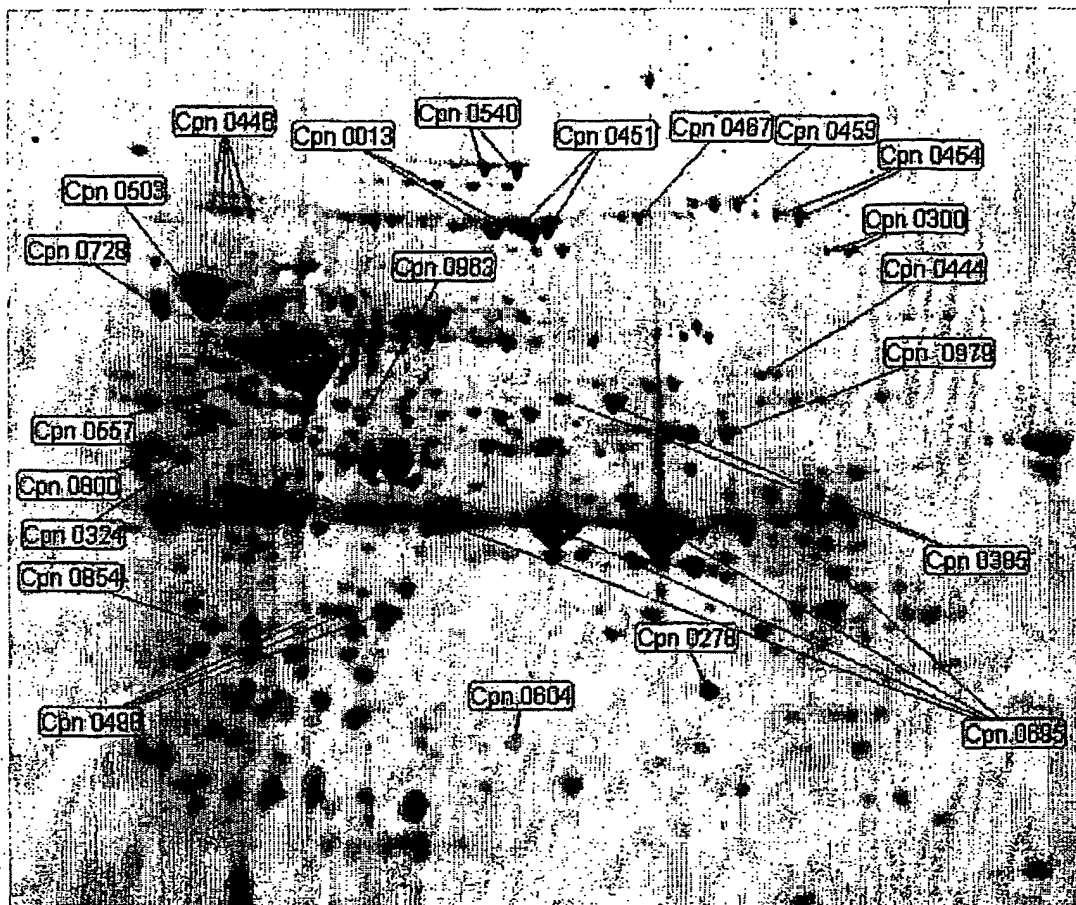


**FIG. 189B**





169/169

**FIGURE 190****FIGURE 191**

SVIVG.VSTNSEHRYHAFQYADGQMVLDLGTLCGPESYAQGVSGDCK  
 KVIVG.HSTRDGEYRAFKYVDGRMIDLGTLCGSASFAGVSDGCK  
 KVIVG.RSETYYGEVHAFCHKNGVMSDLGTLCGSYSAAKGVSAATCK  
 KVIVG.WSTTNGETHAFMHKDETMHDLGTLCGGFSVATGVSAATCK  
 TIIIVGSMESTITRKTAVKVVNNVPTYLGTLCGDASTGLYISGDCT

